TRANSACTIONS

OF THE

AMERICAN MATHEMATICAL SOCIETY

EDITED BY

DUNHAM JACKSON

EDWARD KASNER

HOWARD HAWKS MITCHELL

WITH THE COÖPERATION OF

RAYMOND W. BRINK	EDWARD W. CHITTENDEN	JESSE DOUGLAS
WILLIAM C. GRAUSTEIN	OLIVE C. HAZLETT	EINAR HILLE
AUBREY J. KEMPNER	JOHN R. KLINE	CHARLES N. MOORE
GEORGE Y. RAINICH	JOSEPH F. RITT	CAROLINE E. SEELY
FRANCIS R. SHARPE	JACOB D. TAMARKIN	J. H. M. WEDDERBURN

VOLUME 31

1929

PUBLISHED BY THE SOCIETY

MENASHA, WIS., AND NEW YORK

1929

BOSTON UNIVERSITY
COLLEGE OF LIBERAL ARTS
LIBRARY

General
Bound Jan 1930

COMPOSED, PRINTED AND BOUND BY
The Collegiate Press
GEORGE BANTA PUBLISHING COMPANY
MENASHA, WISCONSIN

39442

QA
1
F00!
v.31

TABLE OF CONTENTS

VOLUME 31, 1929

PAGE

ADAMS, C. R., of Providence, R. I. On the linear partial q-difference equation of general type 360

——— Note on integro-q-difference equations 861

ALBERT, A. A., of Chicago, Ill. A determination of all normal division algebras in sixteen units 253

AYRES, W. L., of Austin, Tex. Concerning the arc-curves and basic sets of a continuous curve, second paper 595

BELL, E. T., of Pasadena, Calif. Certain invariant sequences of polynomials 405

CALUGARÉANO, G., of Paris, France. Les fonctions polygènes comme intégrales d'équations différentielles 372

CAMP, B. H., of Middletown, Conn. The multinomial solid and the chi test 133

CARPENTER, A. F., of Seattle, Wash. Tetrads of ruled surfaces . . . 523

CARSON, J. R., of New York, N. Y. The asymptotic solution of an operational equation 782

CHITTENDEN, E. W., of Iowa City, Iowa. On general topology and the relation of the properties of the class of all continuous functions to the properties of space 290

DICKSON, L. E., of Chicago, Ill. Universal quadratic forms 164

EMCH, A., of Urbana, Ill. Triple and multiple systems, their geometric configurations and groups. 25

FORD, L. R., of Houston, Tex. The limit points of a group 821

FORT, T., of Bethlehem, Pa. Generalized factorial series 233

GEHMAN, H. M., of New Haven, Conn. On extending a continuous (1-1) correspondence (second paper) 241

GROVE, V. G., of East Lansing, Mich. Nets with equal W invariants 845

——— Contributions to the general theory of transformations of nets 853

HAZLETT, O. C., of Urbana, Ill. Homogeneous polynomials with a multiplication theorem 223

HEINEMAN, E. R., of Lubbock, Tex. Generalized Vandermonde determinants 464

HICKSON, A. O., of Providence, R. I. An application of the calculus of variations to boundary value problems 563

HUTCHINSON, J. I., of Ithaca, N. Y. Properties of functions represented by the Dirichlet series $\sum(a\nu+b)^{-s}$, or by linear combinations of such series 322

JACKSON, D., of Minneapolis, Minn. A note on closest approximation 215

KOOPMAN, B. O., of New York, N. Y. On the singularities of linear partial differential equations 687

LANGER, R. E., of Madison, Wis. The boundary problem associated with a differential equation in which the coefficient of the parameter changes sign 1

——— The asymptotic location of the roots of a certain transcendental equation. 837

——— The expansion problem in the theory of ordinary linear differential systems of the second order 868

PAGE

LEHMER, D. N., of Berkeley, Calif. On the congruences connected with certain magic squares 529

LUBBEN, R. G., of Austin, Tex. Separation theorems with applications to questions concerning accessibility and plane continua 503

MCCOY, N. H., of Iowa City, Ia. On commutation formulas in the algebra of quantum mechanics 793

MACDUFFEE, C. C., of Columbus, Ohio. An introduction to the theory of ideals in linear associative algebras 71

MANNING, W. A., of Stanford University, Calif. The degree and class of multiply transitive groups, II 643

MILES, E. R. C., of Houston, Tex. Boundary value problems for potentials of a single layer 190

MILNE, W. E., of Stanford University, Calif. On the degree of convergence of the Gram-Charlier series 422

——— On the degree of convergence of expansions in an infinite interval 907

MORSE, M., of Cambridge, Mass. The foundations of the calculus of variations in the large in m-space (first paper) 379

PIERPONT, J., of New Haven, Conn. Foucault's pendulum in elliptic space 444

RANUM, A., of Ithaca, N. Y. The singular points of analytic space-curves 145, 931

RITT, J. F., of New York, N. Y. Algebraic combinations of exponentials 654

——— On the zeros of exponential polynomials 680

ROOS, C. F., of Chicago, Ill. A general problem of minimizing an integral with discontinuous integrand 58

ROWE, C. H., of Dublin, Ireland. A kinematical treatment of some theorems on normal rectilinear congruences 919

RUPP, C. A., of State College, Pa. An extension of Pascal's theorem. . . 580

RUTLEDGE, G., of Cambridge, Mass. The basic power series of interpolation 807

SHEFFER, I. M., of Chicago, Ill. Expansions in generalized Appell polynomials, and a class of related linear functional equations 261

——— Systems of infinitely many linear differential equations of infinite order, with constant coefficients 281

SHOHAT, J. A. (CHOKHATE, J. A.), of Ann Arbor, Mich. On a certain formula of mechanical quadratures with non-equidistant ordinates . . 448

SUSCHKEWITSCH, A., of Voronesh, Russia. On a generalization of the associative law 204

VANDIVER, H. S., of Austin, Tex. On Fermat's last theorem 613

WALL, H. S., of Madison, Wis. On the Padé approximants associated with the continued fraction and series of Stieltjes 91

——— On extended Stieltjes series 771

WALSH, J. L., of Cambridge, Mass. Note on the expansion of analytic functions in series of polynomials and in series of other analytic functions 53

——— On approximation by rational functions to an arbitrary function of a complex variable. 477

PAGE

WEATHERBURN, C. E., of Christchurch, New Zealand. On curvilinear congruences 117

WIDDER, D. V., of Minneapolis, Minn. On the expansion of analytic functions of the complex variable in generalized Taylor's series 43

——— A generalization of Dirichlet's series and of Laplace's integrals by means of a Stieltjes integral 694

WILDER, R. L., of Ann Arbor, Mich. Concerning zero-dimensional sets in euclidean space 345

WILSON, W. A., of New Haven, Conn. Certain problems related to the cutting of a simply connected plane region by a continuum 552

——— On linear upper semi-continuous collections of bounded continua 829

ZIPPIN, L., of Philadelphia, Pa. A study of continuous curves and their relation to the Janiszewski-Mullikin theorem 744

Errata, volume 31 931

THE BOUNDARY PROBLEM ASSOCIATED WITH A DIFFERENTIAL EQUATION IN WHICH THE COEFFICIENT OF THE PARAMETER CHANGES SIGN*

BY

RUDOLPH E. LANGER

1. Introduction. The boundary value and expansion problem associated with the ordinary linear differential system

$$u''(x) + p(x)u'(x) + [q(x) + \lambda k(x)]u(x) = 0,$$
$$L_j(u) = 0, \qquad j = 1,2,$$

with coefficients real and continuous, and λ a complex parameter, has been extensively studied. The results of these investigations (and of those dealing with the generalized system of order n) comprise an extensive theory for the case in which the coefficient of the parameter maintains its sign in the interval determined by the boundary conditions. Not merely the existence, but the asymptotic forms of the characteristic values and functions are known, and the expansibility of a function, arbitrary within wide restrictions, in a series of characteristic functions has been established.

Far less has been done with the cases in which $k(x)$ either changes sign or otherwise vanishes in the given interval. The existence of infinitely many characteristic values has been variously established,† but their distribution and the form of the corresponding solutions has not been determined. If $k(x)$ changes sign a finite number of times Hilbert's theory of the polar integral equation is applicable and yields the theorem that a function which is continuous together with its first four derivatives and which satisfies the boundary and certain auxiliary conditions is expansible in a uniformly convergent series of characteristic functions. The method of infinitely many variables has been applied with considerable success to the more general case in which $k(x)$ is less restricted.‡ Still the expansion theorems obtained are hardly comparable with those which have been obtained under the hypothesis that the sign of $k(x)$ does not change.

* Presented to the Society, September 8, 1927; received by the editors January 17, 1928.

† For references see the introduction to the paper by L. Lichtenstein, *Zur Analysis der unendlichvielen Variabeln*, Rendiconti del Circolo Matematico di Palermo, vol. 38 (1914), p. 113.

‡ Lichtenstein, loc. cit., also Anna Pell Wheeler, *Linear ordinary self-adjoint differential equations of the second order*, American Journal of Mathematics, vol. 46 (1927), p. 309.

BOSTON UNIVERSITY
COLLEGE OF LIBERAL ARTS
LIBRARY

The present paper is devoted to a discussion of the case in which the coefficient $k(x)$ is the function x^ν, where ν is any real positive constant and the interval includes the point $x=0$. If ν is integral this is a specialization of the equation discussed above since the coefficient remains real. If the integer is even or $x=0$ is an end point of the interval the coefficient merely vanishes, while if ν is an odd integer a change of sign takes place. If ν is not integral the coefficient clearly changes from real to complex.

The method used is that of asymptotic forms. The equation is solved in terms of Bessel's functions and the asymptotic forms of the characteristic values and of the characteristic functions are deduced. It is shown that if $f(x)$ is any integrable function then the expansion in terms of the characteristic functions converges except at the end points to the average $\frac{1}{2}[f(x+)+f(x-)]$ in any interval in which $f(x)$ is of bounded variation and $x\neq 0$. If $f(x)$ is of bounded variation throughout the whole interval the expansion converges also at $x=0$, to the average of the functional values if ν is an integer and otherwise to a weighted average.

2. **The solution of the equation.** We consider the equation

$$u'' + \lambda x^\nu u(x) = 0, \tag{1}$$

in which ν is any real positive constant and λ is a complex parameter. When $x<0$ we specify the symbol x^ν to designate $|x|^\nu e^{\nu\pi i}$. Similarly $\rho^{1/2}$ shall designate $|\rho|^{1/2}e^{(i/2)\arg\rho}$.

It is convenient to set

$$\nu = 2k-2, \quad \lambda = k^2\rho^2, \tag{2}$$

and to consider ρ as a new parameter. If the variables are then changed by the substitutions

$$y(x) = x^{-1/2}u(x), \quad t = \rho x^k,$$

the equation takes the form of the Bessel's equation with solutions

$$y_j(t) = H^{(j)}_{1/(2k)}(t),^* \qquad j = 1,2.$$

Hence we may choose as a set of independent solutions of the equation (1)

$$u_j(x,\rho) = \left(\frac{\pi\rho}{2}\right)^{1/2} e^{\pm(k+1)\pi i/(4k)} x^{1/2} H^{(j)}_{1/(2k)}(\rho x^k), \qquad j = 1,2.\dagger \tag{3}$$

---

* The choice of the Bessel's functions of the third kind is made because of their suitability for the subsequent developments. Cf. Watson, *Theory of Bessel's Functions*, Cambridge University Press, 1922, p. 73. This reference will be indicated in the text by W.

† In combining two formulas by the use of double signs we shall agree to associate the upper signs with $j=1$ and the lower signs with $j=2$.

With the use of the formula [W, p. 74]

$$\frac{d}{dt}\{t^r H_r^{(j)}(t)\} = t^r H_{r-1}^{(j)}(t),$$

we obtain from (3)

$$(3a)\quad u_j'(x,\rho) = k\left(\frac{\pi}{2}\right)^{1/2} e^{\pm(k+1)\pi i/(4k)}\rho^{3/2}x^{k-1/2}H_{1/(2k)-1}^{(j)}(\rho x^k), \qquad j = 1,2.$$

3. **The asymptotic forms of the solutions.** If we adopt as generic symbols $B(t)$ to designate a bounded function, and $I_p(t, r)$ to designate a function of the form

$$I_p(t,r) = \sum_{m=0}^{p-1}\frac{(r,m)}{(-2it)^m} + \frac{B(t)}{t^p},$$

where

$$(r,m) = \frac{\Gamma(r+m+\tfrac{1}{2})}{m!\Gamma(r-m+\tfrac{1}{2})},$$

the asymptotic expressions

$$H_r^{(j)}(t) = \left(\frac{2}{\pi t}\right)^{1/2} e^{\pm i(t-r\pi/2-\pi/4)}I_p(\pm t, r)$$

are available [W, p. 198] and are valid for any integer $p>0$, and uniformly so for $-\pi+\epsilon \leqq \arg t \leqq \pi-\epsilon$, where ϵ is an arbitrarily small positive constant. We find from this that with the use of the abbreviations

$$(4)\quad I(t) \equiv I_p\left(t, \frac{1}{2k}\right),$$

$$\theta_j(x,\rho) \equiv x^{-p/4}e^{\pm i\rho x^k}I(\pm\rho x^k), \qquad j = 1,2,$$

we may write

$$(5)\quad u_j(x,\rho) = \theta_j(x,\rho),$$

for

$$x > 0, \quad -\pi+\epsilon \leqq \arg\rho \leqq \pi-\epsilon.$$

Similarly we deduce as the formula for the derivative

$$(5a)\quad u_j'(x,\rho) = \pm i\rho k x^{p/4}e^{\pm i\rho x^k}I_p\left(t, \frac{1}{2k}-1\right).$$

To obtain the asymptotic forms when $x<0$ we employ the formulas [W, p. 75]

$$(6)\quad H_r^{(j)}(te^{m\pi i}) = \frac{\sin(1\mp m)r\pi}{\sin r\pi}H_r^{(j)}(t) \mp e^{\mp r\pi i}\frac{\sin mr\pi}{\sin r\pi}H_r^{(3-j)}(t),$$

The present paper is devoted to a discussion of the case in which the coefficient $k(x)$ is the function x^ν, where ν is any real positive constant and the interval includes the point $x=0$. If ν is integral this is a specialization of the equation discussed above since the coefficient remains real. If the integer is even or $x=0$ is an end point of the interval the coefficient merely vanishes, while if ν is an odd integer a change of sign takes place. If ν is not integral the coefficient clearly changes from real to complex.

The method used is that of asymptotic forms. The equation is solved in terms of Bessel's functions and the asymptotic forms of the characteristic values and of the characteristic functions are deduced. It is shown that if $f(x)$ is any integrable function then the expansion in terms of the characteristic functions converges except at the end points to the average $\frac{1}{2}[f(x+)+f(x-)]$ in any interval in which $f(x)$ is of bounded variation and $x\neq 0$. If $f(x)$ is of bounded variation throughout the whole interval the expansion converges also at $x=0$, to the average of the functional values if ν is an integer and otherwise to a weighted average.

2. **The solution of the equation.** We consider the equation

$$u'' + \lambda x^\nu u(x) = 0, \tag{1}$$

in which ν is any real positive constant and λ is a complex parameter. When $x<0$ we specify the symbol x^ν to designate $|x|^\nu e^{\nu\pi i}$. Similarly $\rho^{1/2}$ shall designate $|\rho|^{1/2}e^{(i/2)\arg\rho}$.

It is convenient to set

$$\nu = 2k-2, \quad \lambda = k^2\rho^2, \tag{2}$$

and to consider ρ as a new parameter. If the variables are then changed by the substitutions

$$y(x) = x^{-1/2}u(x), \quad t = \rho x^k,$$

the equation takes the form of the Bessel's equation with solutions

$$y_j(t) = H^{(j)}_{1/(2k)}(t),^* \qquad j = 1,2.$$

Hence we may choose as a set of independent solutions of the equation (1)

$$u_j(x,\rho) = \left(\frac{\pi\rho}{2}\right)^{1/2} e^{\pm(k+1)\pi i/(4k)} x^{1/2} H^{(j)}_{1/(2k)}(\rho x^k), \qquad j = 1,2.\dagger \tag{3}$$

* The choice of the Bessel's functions of the third kind is made because of their suitability for the subsequent developments. Cf. Watson, *Theory of Bessel's Functions*, Cambridge University Press, 1922, p. 73. This reference will be indicated in the text by W.

† In combining two formulas by the use of double signs we shall agree to associate the upper signs with $j=1$ and the lower signs with $j=2$.

With the use of the formula [W, p. 74]

$$\frac{d}{dt}\{t^r H_r^{(j)}(t)\} = t^r H_{r-1}^{(j)}(t),$$

we obtain from (3)

(3a) $$u_j'(x,\rho) = k\left(\frac{\pi}{2}\right)^{1/2} e^{\pm(k+1)\pi i/(4k)}\rho^{3/2}x^{k-1/2}H^{(j)}_{1/(2k)-1}(\rho x^k), \qquad j = 1,2.$$

3. **The asymptotic forms of the solutions.** If we adopt as generic symbols $B(t)$ to designate a bounded function, and $I_p(t, r)$ to designate a function of the form

$$I_p(t,r) = \sum_{m=0}^{p-1}\frac{(r,m)}{(-2it)^m} + \frac{B(t)}{t^p},$$

where

$$(r,m) = \frac{\Gamma(r+m+\tfrac{1}{2})}{m!\Gamma(r-m+\tfrac{1}{2})},$$

the asymptotic expressions

$$H_r^{(j)}(t) = \left(\frac{2}{\pi t}\right)^{1/2} e^{\pm i(t-r\pi/2-\pi/4)}I_p(\pm t, r)$$

are available [W, p. 198] and are valid for any integer $p>0$, and uniformly so for $-\pi+\epsilon \leqq \arg t \leqq \pi-\epsilon$, where ϵ is an arbitrarily small positive constant. We find from this that with the use of the abbreviations

(4) $$I(t) \equiv I_p\left(t, \frac{1}{2k}\right),$$

$$\theta_j(x,\rho) \equiv x^{-r/4}e^{\pm i\rho x^k}I(\pm\rho x^k), \qquad j = 1,2,$$

we may write

(5) $$u_j(x,\rho) = \theta_j(x,\rho),$$

for

$$x > 0, \quad -\pi+\epsilon \leqq \arg\rho \leqq \pi-\epsilon.$$

Similarly we deduce as the formula for the derivative

(5a) $$u_j'(x,\rho) = \pm\, i\rho k x^{r/4}e^{\pm i\rho x^k}I_p\left(t, \frac{1}{2k}-1\right).$$

To obtain the asymptotic forms when $x<0$ we employ the formulas [W, p. 75]

(6) $$H_r^{(j)}(te^{m\pi i}) = \frac{\sin(1\mp m)r\pi}{\sin r\pi}H_r^{(j)}(t) \mp e^{\mp r\pi i}\frac{\sin mr\pi}{\sin r\pi}H_r^{(3-j)}(t),$$

The present paper is devoted to a discussion of the case in which the coefficient $k(x)$ is the function x^ν, where ν is any real positive constant and the interval includes the point $x=0$. If ν is integral this is a specialization of the equation discussed above since the coefficient remains real. If the integer is even or $x=0$ is an end point of the interval the coefficient merely vanishes, while if ν is an odd integer a change of sign takes place. If ν is not integral the coefficient clearly changes from real to complex.

The method used is that of asymptotic forms. The equation is solved in terms of Bessel's functions and the asymptotic forms of the characteristic values and of the characteristic functions are deduced. It is shown that if $f(x)$ is any integrable function then the expansion in terms of the characteristic functions converges except at the end points to the average $\frac{1}{2}[f(x+)+f(x-)]$ in any interval in which $f(x)$ is of bounded variation and $x\neq 0$. If $f(x)$ is of bounded variation throughout the whole interval the expansion converges also at $x=0$, to the average of the functional values if ν is an integer and otherwise to a weighted average.

2. **The solution of the equation.** We consider the equation

$$u'' + \lambda x^\nu u(x) = 0, \tag{1}$$

in which ν is any real positive constant and λ is a complex parameter. When $x<0$ we specify the symbol x^ν to designate $|x|^\nu e^{\nu\pi i}$. Similarly $\rho^{1/2}$ shall designate $|\rho|^{1/2}e^{(i/2)\arg\rho}$.

It is convenient to set

$$\nu = 2k-2, \quad \lambda = k^2\rho^2, \tag{2}$$

and to consider ρ as a new parameter. If the variables are then changed by the substitutions

$$y(x) = x^{-1/2}u(x), \quad t = \rho x^k,$$

the equation takes the form of the Bessel's equation with solutions

$$y_j(t) = H^{(j)}_{1/(2k)}(t),^* \qquad j = 1,2.$$

Hence we may choose as a set of independent solutions of the equation (1)

$$u_j(x,\rho) = \left(\frac{\pi\rho}{2}\right)^{1/2} e^{\pm(k+1)\pi i/(4k)} x^{1/2} H^{(j)}_{1/(2k)}(\rho x^k), \qquad j = 1,2.\dagger \tag{3}$$

* The choice of the Bessel's functions of the third kind is made because of their suitability for the subsequent developments. Cf. Watson, *Theory of Bessel's Functions*, Cambridge University Press, 1922, p. 73. This reference will be indicated in the text by W.

† In combining two formulas by the use of double signs we shall agree to associate the upper signs with $j=1$ and the lower signs with $j=2$.

With the use of the formula [W, p. 74]

$$\frac{d}{dt}\{t^r H_r^{(j)}(t)\} = t^r H_{r-1}^{(j)}(t),$$

we obtain from (3)

$$(3a)\quad u_j'(x,\rho) = k\left(\frac{\pi}{2}\right)^{1/2} e^{\pm(k+1)\pi i/(4k)}\rho^{3/2}x^{k-1/2}H^{(j)}_{1/(2k)-1}(\rho x^k), \qquad j = 1,2.$$

3. **The asymptotic forms of the solutions.** If we adopt as generic symbols $B(t)$ to designate a bounded function, and $I_p(t, r)$ to designate a function of the form

$$I_p(t,r) = \sum_{m=0}^{p-1}\frac{(r,m)}{(-2it)^m} + \frac{B(t)}{t^p},$$

where

$$(r,m) = \frac{\Gamma(r+m+\tfrac{1}{2})}{m!\Gamma(r-m+\tfrac{1}{2})},$$

the asymptotic expressions

$$H_r^{(j)}(t) = \left(\frac{2}{\pi t}\right)^{1/2} e^{\pm i(t-r\pi/2-\pi/4)}I_p(\pm t, r)$$

are available [W, p. 198] and are valid for any integer $p>0$, and uniformly so for $-\pi+\epsilon \leqq \arg t \leqq \pi-\epsilon$, where ϵ is an arbitrarily small positive constant. We find from this that with the use of the abbreviations

$$(4)\quad I(t) \equiv I_p\left(t, \frac{1}{2k}\right),$$

$$\theta_j(x,\rho) \equiv x^{-r/4}e^{\pm i\rho x^k}I(\pm\rho x^k), \qquad j = 1,2,$$

we may write

$$(5)\quad u_j(x,\rho) = \theta_j(x,\rho),$$

for

$$x > 0, \quad -\pi+\epsilon \leqq \arg\rho \leqq \pi-\epsilon.$$

Similarly we deduce as the formula for the derivative

$$(5a)\quad u_j'(x,\rho) = \pm i\rho k x^{r/4}e^{\pm i\rho x^k}I_p\left(t, \frac{1}{2k}-1\right).$$

To obtain the asymptotic forms when $x<0$ we employ the formulas [W, p. 75]

$$(6)\quad H_r^{(j)}(te^{m\pi i}) = \frac{\sin(1\mp m)r\pi}{\sin r\pi}H_r^{(j)}(t) \mp e^{\mp r\pi i}\frac{\sin mr\pi}{\sin r\pi}H_r^{(3-j)}(t),$$

which are valid when m is an integer. Setting $|x|=\xi$, i.e. $x=\xi e^{\nu i}$, and defining the integer K and the fractions η, η' by the relations

$$K-1<k\leqq K,$$
$$k=K-\eta=(K-1)+\eta', \tag{7}$$

we find from (3) and (6) that if

$$\rho'=\rho e^{-\eta\pi i} \tag{8}$$

and arg ρ' is restricted to the interval $(-\pi+\epsilon,\pi-\epsilon)$, then

$$u_j(x,\rho)=i^{1+\eta}[c_{j1}\theta_1(\xi,\rho')+c_{j2}\theta_2(\xi,\rho')], \qquad j=1,2, \tag{9}$$

where

$$c_{11}=-\cos\frac{\eta'\pi}{2k}\csc\frac{\pi}{2k},\quad c_{12}=c_{21}=-i\cos\frac{\eta\pi}{2k}\csc\frac{\pi}{2k}, \tag{10}$$
$$c_{22}=\cos\frac{(1+\eta)\pi}{2k}\csc\frac{\pi}{2k}.$$

In terms of x formula (9) is

$$u_j(x,\rho)=i^K x^{-\nu/4}[c_{j1}\,e^{i\rho x^k}I(\rho x^k)+c_{j,3-l}e^{-i\rho x^k}I(-\rho x^k)], \qquad j=1,2, \tag{11}$$

for

$$x<0,\quad (-1+\eta)\pi+\epsilon\leqq\arg\rho\leqq(1+\eta)\pi-\epsilon,$$

where $l=1$ if K is even, and $l=2$ if K is odd.

In (11) as in (5), (5a) the asymptotic form of u_j' may be obtained from that of u_j by differentiating only the exponential factors and replacing $I(t)$ by $I_\nu(t,\,1/(2k)-1)$.

When ν is an integer it follows from the work of Birkhoff* that for ρ confined to any quadrant formed by the axes of reals and imaginaries every solution of equation (1) is asymptotically of the form

$$u(x)=x^{-\nu/4}[a_1e^{i\rho x^k}v_1(x,\rho)+a_2e^{-i\rho x^k}v_2(x,\rho)],$$

for $x\geqq\delta>0$, while for $x\leqq-\delta$ such a solution is of similar structure with a_1 and a_2 replaced by different constants $\bar a_1$ and $\bar a_2$ respectively. From formulas (5) and (11) the relation between these sets of constants for the quadrants in which (5) and (11) are both valid is found to be

$$\bar a_1=i^K(c_{1,l}a_1+c_{2,l}a_2),$$
$$\bar a_2=i^K(c_{1,3-l}a_1+c_{2,3-l}a_2).$$

* Birkhoff, *Asymptotic solutions of differential equations*, these Transactions, vol. 9 (1908), pp. 219–231.

The asymptotic expressions obtained display the functional form of the solutions when $|\rho x^k|$ is sufficiently large. For small values of $|\rho x^k|$ we observe that since

$$t^{\mu}H^{(j)}_{\pm\mu}(t), \qquad j = 1,2, \quad \mu > 0,$$

is bounded when t is bounded, it follows from (3) and (3a) that

$$\begin{aligned} &|\, x^{\nu/4}u_j(x,\rho)\,| < M_1, \\ &|\, u_j'(x,\rho)\,| < M_2|\,\rho\,|^{1/2+1/(2k)}, \qquad \text{for } |\,\rho x^k| < M, \end{aligned} \tag{12}$$

Combining these inequalities with formulas (5) and (5a) we find that

$$\begin{aligned} u_j(x,\rho) &= x^{-\nu/4}e^{\pm i\rho x^k}B(x,\rho), \\ u_j'(x,\rho) &= \pm\, ik\rho e^{\pm i\rho x^k}B(x,\rho), \qquad \text{for } 0 \leqq x. \end{aligned} \tag{13}$$

Lastly we obtain from the formulas [W, pp. 74, 40]

$$\lim_{t\to 0} t^{\mu}H^{(j)}_{\mu}(t) = A_j, \qquad \lim_{t\to 0} t^{\mu}H^{(j)}_{-\mu}(t) = A_j e^{\pm\mu\pi i}, \qquad \mu > 0,$$

where

$$A_j = \frac{\pm\, 2^{\mu}}{i\Gamma(1-\mu)\sin\mu\pi}$$

and

$$\Gamma(z)\Gamma(1-z) = \frac{\pi}{\sin\pi z}, \tag{14}$$

and the relations (3) and (3a)

$$\begin{aligned} u_j(0,\rho) &= \frac{\mp\, i}{\pi^{1/2}}e^{\pm(k+1)\pi i/(4k)}\Gamma\left(\frac{1}{2k}\right)\left(\frac{\rho}{2}\right)^{1/2-1/(2k)}, \\ u_j'(0,\rho) &= \frac{\pm\, 2ki}{\pi^{1/2}}e^{\pm(k-1)\pi i/(4k)}\Gamma\left(1-\frac{1}{2k}\right)\left(\frac{\rho}{2}\right)^{1/2+1/(2k)}, \quad j = 1,2. \end{aligned} \tag{15}$$

4. **The boundary problem. The characteristic values.** We consider now the boundary problem obtained by adjoining to equation (1) the boundary conditions

$$u(\beta) = 0, \qquad u(-\alpha) = 0,^{*} \qquad \alpha > 0, \qquad \beta > 0. \tag{16}$$

Since every solution of (1) may be written in the form

$$u(x) = a_1u_1(x) + a_2u_2(x) \tag{17}$$

---

* For brevity we shall use the notation $u(x)$ in place of $u(x, \rho)$ when no confusion is apt to result.

which are valid when m is an integer. Setting $|x|=\xi$, i.e. $x=\xi e^{\nu i}$, and defining the integer K and the fractions η, η' by the relations

$$K-1<k\leqq K,\qquad k=K-\eta=(K-1)+\eta', \tag{7}$$

we find from (3) and (6) that if

$$\rho'=\rho e^{-\eta\pi i} \tag{8}$$

and arg ρ' is restricted to the interval $(-\pi+\epsilon,\pi-\epsilon)$, then

$$u_j(x,\rho)=i^{1+\nu}[c_{j1}\theta_1(\xi,\rho')+c_{j2}\theta_2(\xi,\rho')], \qquad j=1,2, \tag{9}$$

where

$$c_{11}=-\cos\frac{\eta'\pi}{2k}\csc\frac{\pi}{2k},\quad c_{12}=c_{21}=-i\cos\frac{\eta\pi}{2k}\csc\frac{\pi}{2k}, \tag{10}$$

$$c_{22}=\cos\frac{(1+\eta)\pi}{2k}\csc\frac{\pi}{2k}.$$

In terms of x formula (9) is

$$u_j(x,\rho)=i^K x^{-\nu/4}[c_{j1}e^{i\rho z^k}I(\rho x^k)+c_{j,3-l}e^{-i\rho z^k}I(-\rho x^k)], \qquad j=1,2, \tag{11}$$

for

$$x<0,\quad(-1+\eta)\pi+\epsilon\leqq\arg\rho\leqq(1+\eta)\pi-\epsilon,$$

where $l=1$ if K is even, and $l=2$ if K is odd.

In (11) as in (5), (5a) the asymptotic form of u_j' may be obtained from that of u_j by differentiating only the exponential factors and replacing $I(t)$ by $I_\nu(t,1/(2k)-1)$.

When ν is an integer it follows from the work of Birkhoff* that for ρ confined to any quadrant formed by the axes of reals and imaginaries every solution of equation (1) is asymptotically of the form

$$u(x)=x^{-\nu/4}[a_1e^{i\rho z^k}v_1(x,\rho)+a_2e^{-i\rho z^k}v_2(x,\rho)],$$

for $x\geqq\delta>0$, while for $x\leqq-\delta$ such a solution is of similar structure with a_1 and a_2 replaced by different constants $\bar{a}_1$ and $\bar{a}_2$ respectively. From formulas (5) and (11) the relation between these sets of constants for the quadrants in which (5) and (11) are both valid is found to be

$$\bar{a}_1=i^K(c_{1,l}a_1+c_{2,l}a_2),$$
$$\bar{a}_2=i^K(c_{1,3-l}a_1+c_{2,3-l}a_2).$$

* Birkhoff, *Asymptotic solutions of differential equations*, these Transactions, vol. 9 (1908), pp. 219-231.

The asymptotic expressions obtained display the functional form of the solutions when $|\rho x^k|$ is sufficiently large. For small values of $|\rho x^k|$ we observe that since

$$t^{\mu}H^{(j)}_{\pm\mu}(t), \qquad j = 1,2, \quad \mu > 0,$$

is bounded when t is bounded, it follows from (3) and (3a) that

$$\begin{aligned} &|\, x^{\nu/4}u_j(x,\rho)\,| < M_1, \\ &|\, u_j'(x,\rho)\,| < M_2|\,\rho\,|^{1/2+1/(2k)}, \qquad \text{for } |\,\rho x^k| < M. \end{aligned} \tag{12}$$

Combining these inequalities with formulas (5) and (5a) we find that

$$\begin{aligned} u_j(x,\rho) &= x^{-\nu/4}e^{\pm i\rho x^k}B(x,\rho), \\ u_j'(x,\rho) &= \pm\, ik\rho e^{\pm i\rho x^k}B(x,\rho), \qquad \text{for } 0 \leqq x. \end{aligned} \tag{13}$$

Lastly we obtain from the formulas [W, pp. 74, 40]

$$\lim_{t\to 0} t^{\mu}H^{(j)}_{\mu}(t) = A_j, \qquad \lim_{t\to 0} t^{\mu}H^{(j)}_{-\mu}(t) = A_j e^{\pm\mu\pi i}, \qquad \mu > 0,$$

where

$$A_j = \frac{\pm\, 2^{\mu}}{i\Gamma(1-\mu)\sin\mu\pi}$$

and

$$\Gamma(z)\Gamma(1-z) = \frac{\pi}{\sin\pi z}, \tag{14}$$

and the relations (3) and (3a)

$$\begin{aligned} u_j(0,\rho) &= \frac{\mp\, i}{\pi^{1/2}}e^{\pm(k+1)\pi i/(4k)}\Gamma\left(\frac{1}{2k}\right)\left(\frac{\rho}{2}\right)^{1/2-1/(2k)}, \\ u_j'(0,\rho) &= \frac{\pm\, 2ki}{\pi^{1/2}}e^{\pm(k-1)\pi i/(4k)}\Gamma\left(1-\frac{1}{2k}\right)\left(\frac{\rho}{2}\right)^{1/2+1/(2k)}, \qquad j = 1,2. \end{aligned} \tag{15}$$

4. **The boundary problem. The characteristic values.** We consider now the boundary problem obtained by adjoining to equation (1) the boundary conditions

$$u(\beta) = 0, \qquad u(-\alpha) = 0,^{*} \qquad \alpha > 0, \qquad \beta > 0. \tag{16}$$

Since every solution of (1) may be written in the form

$$u(x) = a_1u_1(x) + a_2u_2(x) \tag{17}$$

---

* For brevity we shall use the notation $u(x)$ in place of $u(x, \rho)$ when no confusion is apt to result.

it is clear that there exists a solution satisfying conditions (16) if and only if the value of ρ is such that

$$\Omega(\beta, -\alpha) = 0, \tag{18}$$

where

$$\Omega(s,t) = \begin{vmatrix} u_1(s) & u_2(s) \\ u_1(t) & u_2(t) \end{vmatrix}.$$

Since by (2) any half ρ-plane maps into the entire λ-plane we may restrict ourselves, in seeking the roots of equation (18), to any half-plane

$$-\delta \leqq \arg \rho \leqq \pi - \delta, \text{ where } 0 < \delta < \eta'\pi. \tag{19}$$

For such values the asymptotic forms of $u_j(x, \rho)$ and of $u_j(x, \rho')$ are given by formulas (5) when $x>0$. We must distinguish between the cases in which ν is or is not an even integer.

Case 1. *ν not an even integer.* In this case the rays

$$\arg \rho = 0, \tag{20}$$

$$\arg \rho' = 0, \quad \text{i.e.} \quad \arg \rho = \eta\pi, \tag{21}$$

are distinct. Suppose now that ρ is confined to a sector S_1 of the half-plane (19) which includes the ray (20) but excludes the ray (21). In such a sector $R(i\rho')>0$, and from (4) $\theta_2(x, \rho')$ is asymptotically negligible in comparison with $\theta_1(x, \rho')$ for $x>0$. Hence we find in this sector

$$\Omega(\beta, -\alpha) \sim i^{1+\eta} \begin{vmatrix} \theta_1(\beta,\rho) & \theta_2(\beta,\rho) \\ c_{11}\theta_1(\alpha,\rho') & c_{21}\theta_1(\alpha,\rho') \end{vmatrix}^*,$$

and the equation (18) takes the form

$$c_{21}e^{i\rho\beta^k}I(\rho\beta^k) - c_{11}e^{-i\rho\beta}I(-\rho\beta^k) = 0. \tag{22}$$

If we set

$$r_{1,m} = \beta^{-k}\left[\left(m - \frac{1}{4}\right)\pi + \frac{1}{2i}\log\left(\cos\frac{\eta'\pi}{2k}\sec\frac{\eta\pi}{2k}\right)\right], \tag{23}$$

and observe that

$$\log I(-\rho\beta^k) - \log I(\rho\beta^k) = \frac{\left(\frac{1}{2k}, 1\right)}{i\rho\beta^k} + \frac{B(\rho)}{\rho^3},$$

* We use the symbol $\sim$ to indicate that the functions are of the same asymptotic form. Thus if in this formula the sign $\sim$ were to be replaced by $=$ it would be necessary to replace in the right hand member the quantity $i^{1+\eta}\theta_2(\alpha\rho')[c_{12}\theta_1(\beta)+c_{12}\theta_2(\beta)]$, which has been dropped. By (4) it is evident, however, that this quantity is of lower order of magnitude when $|\rho|$ is large than the error due to stopping at any definite term of the asymptotic expansion of the expression retained.

we may write (22) in the form

$$\rho - \left[r_{1,m} + \frac{\left(\frac{1}{2k}, 1\right)}{2\rho\beta^{2k}} + \frac{B(\rho)}{\rho^3} \right] = 0. \tag{22a}$$

Let the points $r_{1,m}$ in the ρ-plane be surrounded now by small circles with centers at $r_{1,m}$ and radius ϵ. Then for m sufficiently large the parenthesis in (22a) will have a value lying within this small circle for every value of ρ on the circumference, and as ρ traces the circumference the argument of the left hand member of (22a) increases by 2π. Hence (22a) has just one root $\rho_{1,m}$ within the small circle. From (22a) we may write then

$$\rho_{1,m} = r_{1,m} + B(m)/m \cdot$$

Computing the term in $1/\rho$ from this and observing that $r_{1,m}$ is of the order of m we obtain the formula

$$\rho_{1,m} = r_{1,m} + \frac{\left(\frac{1}{2k}, 1\right)}{2r_{1,m}\beta^{2k}} + \frac{B(m)}{m^3} \cdot \tag{24}$$

In similar manner if ρ is restricted to a sector S_2 of the half-plane (19) which includes the ray (21) but excludes the ray (20) we have $R(i\rho) < 0$ and

$$\Omega(\beta, -\alpha) \sim -\theta_2(\beta, \rho)\theta_1(-\alpha, \rho).$$

The characteristic equation (18) now takes the form

$$c_{11}e^{i\rho'\alpha^k}I(\rho\alpha^k) + c_{12}e^{-i\rho'\alpha^k}I(-\rho'\alpha^k) = 0. \tag{25}$$

If we set

$$r_{2,m} = \alpha^{-k}\left[\left(m - \frac{1}{4}\right)\pi - \frac{1}{2i}\log\left(\cos\frac{\eta'\pi}{2k}\sec\frac{\eta\pi}{2k}\right)\right], \tag{26}$$

we find precisely as in the preceding discussion, if we recall (8), that

$$\rho_{2,m} = e^{\eta\pi i}\left[r_{2,m} + \frac{\left(\frac{1}{2k}, 1\right)}{2r_{2,m}\alpha^{2k}} + \frac{B(m)}{m^3} \right]. \tag{27}$$

We may state then the following theorem:

THEOREM 1. *If in equation* (1) ν *is not an even integer the infinitely many characteristic values of* ρ *are asymptotically spaced at regular intervals along two distinct straight lines in the complex* ρ*-plane, the formulas for these values being given by* (24) *and* (27) *above.*

We observe in particular that when ν is an odd integer then $\eta=\eta'=\frac{1}{2}$, and the logarithmic terms in (23) and (26) vanish, so that $\rho_{1,m}$ is asymptotically real and $\rho_{2,m}$ asymptotically pure imaginary.* In this case we find from (2) the theorem

THEOREM 1a. *If in equation* (1) ν *is an odd integer the infinitely many characteristic values are given asymptotically by the formulas*

$$\lambda_{1,m}=k^2\beta^{-2k}\left[\left(m-\frac{1}{4}\right)^2\pi^2+\left(\frac{1}{2k},\,1\right)\right]+\frac{B(m)}{m^2},$$

$$\lambda_{2,m}=-k^2\alpha^{-2k}\left[\left(m-\frac{1}{4}\right)^2\pi^2+\left(\frac{1}{2k},\,1\right)\right]+\frac{B(m)}{m^2}.$$

Case 2. ν *an even integer.* In this case we find from (2), (7), and (9) that k is an integer and hence that $K=k$, $\eta=0$, $\eta'=1$, and $\rho'=\rho$. From the form

$$\Omega(\beta,-\alpha)=i[c_{21}\theta_1(\beta)\theta_1(\alpha)+c_{22}\theta_1(\beta)\theta_2(\alpha)-c_{11}\theta_2(\beta)\theta_1(\alpha)-c_{12}\theta_2(\beta)\theta_2(\alpha)] \tag{28}$$

the equation (18) becomes

$$e^{i\rho(\beta k+\alpha k)}I(\rho\beta^k)I(\rho\alpha^k)+i\left(\cos\frac{\pi}{2k}\right)e^{i\rho(\beta k-\alpha k)}I(\rho\beta^k)I(-\rho\alpha^k)$$
$$+i\left(\cos\frac{\pi}{2k}\right)e^{-i\rho(\beta k-\alpha k)}I(-\rho\beta^k)I(\rho\alpha^k)-e^{-i\rho(\beta k+\alpha k)}I(-\rho\beta^k)I(-\rho\alpha^k)=0. \tag{29}$$

The equation of this type has been studied,† and it is known (i) that its roots are asymptotic to those of the equation

$$e^{2i\rho(\beta k+\alpha k)}+i\left(\cos\frac{\pi}{2k}\right)e^{2i\rho\beta k}+i\left(\cos\frac{\pi}{2k}\right)e^{2i\rho\alpha k}-1=0, \tag{29a}$$

and (ii) that the roots of this latter equation lie in a strip bounded by parallels to the axis of reals,‡ and that for $|\rho|>N$ at most three roots lie between any two lines perpendicular to this axis and less than the distance

$$\frac{2\pi\alpha^k}{\beta^k+\alpha^k}$$

apart.

---

* It is known, of course, that $\rho_{1,m}$ and $\rho_{2,m}$ are actually real and pure imaginary respectively. Cf. Ince, *Ordinary Differential Equations*, London, 1927, p. 238.

† Tamarkin, *Some General Problems of the Theory of Linear Differential Equations* etc. (in Russian), Petrograd, 1917, also C. E. Wilder, these Transactions, vol. 18 (1917), p. 420.

‡ They are actually real. See note * above.

We observe that in the particular cases in which β^k and α^k are commensurable, say

$$\beta^k = (p/q)\alpha^k,$$

where p and q are integers, the solution of equation (29a) is an algebraic problem, the equation being algebraic of the degree $p+q$ in $e^{2i\rho\alpha^k q^{-1}}$. Denoting the roots of this equation by r_j, $j=1, 2, \cdots, p+q$, we have then

$$e^{2i\rho\alpha^k/q} = r_j,$$

from which the characteristic values are found to be

$$\rho_{j,m} = q\alpha^{-k}\left[m\pi + \frac{1}{2i}\log r_j + \epsilon_j(m)\right]$$

where

$$\lim_{m\to\infty} \epsilon_j(m) = 0.$$

We consider more explicitly the particular case in which

$$\alpha = \beta. \tag{30}$$

The determination of the characteristic values is in this case easily possible for the equation (28) reduces to the quadratic

$$\theta_1^2(\beta) + 2i\left(\cos\frac{\pi}{2k}\right)\theta_1(\beta)\theta_2(\beta) - \theta_2^2(\beta) = 0,$$

which we write in the form

$$\{\theta_1(\beta) + ie^{\pi i/(2k)}\theta_2(\beta)\}\{\theta_1(\beta) + ie^{-\pi i/(2k)}\theta_2(\beta)\} = 0. \tag{31}$$

From this the characteristic values are found directly to be

$$\rho_{j,m} = \beta^{-k}\left[\left(m - \frac{1}{4} \pm \frac{1}{4k}\right)\pi - \frac{\left(\frac{1}{2k}, 1\right)}{2\left(m - \frac{1}{4} \pm \frac{1}{4k}\right)\pi}\right] + \frac{B(m)}{m^2}, \quad j=1,2, \tag{32}$$

and we have the following theorem:

THEOREM 2. *If in equation* (1) ν *is an even integer and* $\alpha=\beta$, *the infinitely many characteristic values are given asymptotically by the formula*

$$\lambda_{j,m} = k^2\beta^{-2k}\left[\left(m - \frac{1}{4} \pm \frac{1}{4k}\right)^2\pi^2 - \left(\frac{1}{2k}, 1\right)\right] + \frac{B(m)}{m^2}, \quad j = 1,2. \tag{33}$$

It appears from this that the case ν an even integer is irregular in the

sense that the characteristic values (32) are not the limiting values which would be obtained from (24) and (27) by allowing η to approach zero,i.e. by allowing ν to approach the even integer.

5. **The characteristic functions.** When ρ is a characteristic value the constants a_1, a_2 in (17) may be chosen to make the solution satisfy the boundary conditions. The solution (17) becomes then a characteristic function. We proceed to derive its asymptotic form. For convenience we shall omit the subscripts on ρ, understanding that in the discussion of this section ρ is always taken to be a characteristic value.

Case 1. *ν not an even integer.* Consider first the function corresponding to $\rho=\rho_{1,m}$. Choosing the constants in (17) thus:

$$a_1 = -u_2(\beta,\rho), \qquad a_2 = u_1(\beta,\rho),$$

we find

$$u(x) = \Omega(\beta,x,\rho). \tag{34}$$

The characteristic equation (22) is in this case

$$c_{11}\theta_2(\beta,\rho) - c_{21}\theta_1(\beta,\rho) \sim 0,$$

and substituting from this in (34) we find

$$u(x) \sim -\left(\frac{c_{21}}{c_{11}}\right)^{1/2}\theta_1(\beta,\rho)\left[\left(\frac{c_{21}}{c_{11}}\right)^{1/2}\theta_1(x,\rho) - \left(\frac{c_{11}}{c_{21}}\right)^{1/2}\theta_2(x,\rho)\right]. \tag{35}$$

With the use of the formulas (4) the form of $u(x)$ when $x>0$ is thus obtained.

If $x<0$ we observe that because of (18) we may write

$$\Omega(\beta,x,\rho) = \frac{u_1(\beta,\rho)}{u_1(-\alpha,\rho)}\Omega(-\alpha,x,\rho).$$

Substituting from (9) for $u_j(x)$ and $u_j(-\alpha)$, and observing that

$$c_{11}c_{22} - c_{12}c_{21} = 1, \tag{36}$$

we derive the result

$$u(x,\rho) = \frac{i^{1+\eta}\theta_1(\beta,\rho)}{c_{11}\theta_1(\alpha,\rho') + c_{12}\theta_2(\alpha,\rho')}\Omega(\alpha,\xi,\rho'). \tag{37}$$

Now when $\rho=\rho_{1,m}$, $\theta_2(\alpha,\rho')$ is asymptotically negligible in comparison with $\theta_1(\alpha,\rho')$ and hence we may replace the denominator of (37) by $c_{11}\theta_1(\alpha,\rho')$. Lastly if we remove from (35) and (37) the factor

$$-2i\left(\frac{c_{21}}{c_{11}}\right)^{1/2}\theta_1(\beta,\rho),$$

which is independent of x, we have the following theorem:

THEOREM 3. *If in equation* (1) ν *is not an even integer the characteristic function* $u_{1,m}(x)$ *corresponding to the characteristic value* $\rho_{1,m}$ *is of the asymptotic form*

$$u_{1,m}(x) \sim \frac{x^{-\nu/4}}{2i}\left[\left(\frac{c_{21}}{c_{11}}\right)^{1/2} e^{i\rho x^k} I(\rho x^k) - \left(\frac{c_{11}}{c_{21}}\right)^{1/2} e^{-i\rho x^k} I(-\rho x^k)\right]$$

when $x > 0$,

$$u_{1,m}(x) \sim \frac{\xi^{-\nu/4}}{-2i^{-\eta}(c_{11}c_{21})^{1/2}}\left[e^{-i\rho'\xi^k} I(-\rho'\xi^k) - e^{-i\rho'(2\alpha^k-\xi^k)} \frac{I(-\rho'\alpha^k)}{I(\rho'\alpha^k)} I(\rho'\xi^k)\right]$$

when $x < 0$, $\rho = \rho_{1,m}$.

To derive the form of $u_{2,m}(x)$ we consider the solution

$$u(x) = -i^{-2\eta}\Omega(x, -\alpha, \rho). \tag{38}$$

When $x<0$ we obtain from this by means of the equations (9) and (36)

$$u(x) = \Omega(\xi, \alpha, \rho').$$

The characteristic equation in this case is (25), i.e.

$$c_{11}\theta_1(\alpha,\rho') + c_{12}\theta_2(\alpha,\rho') \sim 0.$$

Observing that (38) and (25) may be obtained from (34) and (22) by interchanging the following letters:

(39) u_1 with u_2, θ_1 with θ_2, c_{21} with $-c_{12}$, β with α, ρ with ρ', x with ξ,

it is clear that the result, namely the form of $u(x)$ when $x<0$, may be similarly obtained from (35).

When $x>0$ we have because of (18) the relations

$$\Omega(x, -\alpha, \rho) = \frac{u_2(-\alpha,\rho)}{u_2(\beta,\rho)}\Omega(x,\beta,\rho) = \frac{i^{1+\eta}\theta_2(\alpha,\rho')}{c_{11}\theta_2(\beta,\rho) - c_{21}\theta_1(\beta,\rho)}\Omega(x,\beta,\rho),$$

and hence from (38)

$$u(x) = -\frac{i^{1-\eta}\theta_2(\alpha,\rho')}{c_{11}\theta_2(\beta,\rho) - c_{21}\theta_1(\beta,\rho)}\Omega(x,\beta,\rho). \tag{40}$$

Observing that the interchanges of letters (39) lead from (37) to (40) except for a factor $-i^{-2\eta}$ it is clear that the result is again derivable from that of the preceding discussion. Removing the factor

$$2i\left(\frac{-c_{12}}{c_{11}}\right)^{1/2}\theta_2(\alpha,\rho')$$

from the solution considered we may state the following theorem:

THEOREM 3_2. *If in equation* (1) ν *is not an even integer the asymptotic forms of the characteristic functions* $u_{2,m}(x)$ *corresponding to* $\rho=\rho_{2,m}$ *are given by the formulas*

$$u_{2,m}(x) \sim \frac{x^{-\nu/4}}{-2i^{\nu}(-c_{11}c_{12})^{1/2}}\left[e^{i\rho x^k}I(\rho x^k) - e^{i\rho(2\beta k - x^k)}\frac{I(\rho\beta^k)}{I(-\rho\beta^k)}I(-\rho x^k)\right]$$

when $x>0$,

$$u_{2,m}(x) \sim \frac{\xi^{-\nu/4}}{2i}\left[\left(\frac{c_{11}}{-c_{12}}\right)^{1/2} e^{i\rho'\xi^k}I(\rho'\xi^k) - \left(\frac{-c_{12}}{c_{11}}\right)^{1/2} e^{-i\rho'\xi^k}I(-\rho'\xi^k)\right]$$

when $x<0$, $\rho=\rho_{2,m}$.

In particular if ν is an odd integer these forms may be somewhat reduced and we have the following theorem:

THEOREM 3a. *If in equation* (1) ν *is an odd integer the characteristic functions* $u_{1,m}(x)$ *corresponding to the characteristic values* $\rho=\rho_{1,m}$ *are given by the formulas*

$$u_{1,m}(x) \sim x^{-\nu/4}\left[\left\{1+\frac{B(mx^k)}{(mx^k)^2}\right\}\sin\left(\rho_{1,m}x^k+\frac{\pi}{4}\right) + \frac{B(m,x^k)}{mx^k}\cos\left(\rho_{1,m}x^k+\frac{\pi}{4}\right)\right],$$

when $x>0$,

$$u_{1,m}(x) \sim -\xi^{-\nu/4}\left(\sin\frac{\pi}{4k}\right)\left[e^{-\rho_{1,m}\xi^k}\left\{1+\frac{B(m\xi^k)}{m\xi^k}\right\} - e^{-\rho_{1,m}(2\alpha k-\xi^k)}\left\{1+\frac{B(m\xi^k)}{m\xi^k}\right\}\right],$$

when $x<0$.

The form of the characteristic function $u_{2,m}(x)$ *for* $x>0$ *(or* $x<0$*) is obtainable from that of* $u_{1,m}(x)$ *for* $x<0$ *(or* $x>0$*) by interchanging* x *with* ξ *and* α *with* β.

From this we see that for large values of m any solution has the character of $|x|^{-\nu/4}\sin(\rho_{1,m}|x|^k+\pi/4)$ in any interval of the x-axis lying entirely on one side of the origin, while in any interval of the x-axis on the other side of the origin the graphs of the solutions lie arbitrarily close to the axis for m sufficiently large.

It is desirable also to note the values taken on by the characteristic functions at $x=0$. Removing the constant factors specified in the discussion above and substituting from (15) we find

$$(41)\qquad u_{j,m}(0) = C_j \rho_{j,m}^{1/2-1/(2k)}, \quad u'_{j,m}(0) = C'_j \rho_{j,m}^{1/2+1/(2k)}, \qquad j = 1,2,$$

where C_j and C_j' are constants independent of m. Since $|\rho_{j,m}|$ is of the order m it is seen that $|u_{j,m}(0)|$, and $|u'_{j,m}(0)|$, become infinite with m.

Case 2. *ν an even integer.* Inasmuch as we have derived the asymptotic forms of the characteristic values when ν is even only for the case $\alpha=\beta$, we shall confine our discussion of the characteristic functions also to this case. Setting

$$(42)\qquad u(x) = \Omega(\beta, x),$$

and observing that from (31) we have

$$\theta_1(\beta,\rho_{j,m}) = -ie^{\pm \pi i/(2k)}\theta_2(\beta,\rho_{j,m}), \qquad j = 1,2,$$

we obtain directly on dropping the factor

$$-2e^{\pi i/4 \pm \pi i/(4k)}\theta_2(\beta,\rho_{j,m})$$

the form

$$(43)\quad u_{j,m}(x) \sim \frac{x^{-\nu/4}}{2i}\left[e^{i(\rho x^k + \pi/4 \mp \pi/(4k))}I(\rho x^k) - e^{-i(\rho x^k+\pi/4\mp\pi/(4k))}I(-\rho x^k)\right]$$

for $x>0$, $\rho=\rho_{1,m}$, or in an alternative form

$$(44)\quad u_{j,m}(x) \sim x^{\nu/4}\left[\left\{1 + \frac{B(mx^k)}{(mx^k)^2}\right\}\sin\left(\rho_{j,m}x^k + \frac{\pi}{4} \mp \frac{\pi}{4k}\right) - \frac{B(mx^k)}{mx^k}\cos\left(\rho_{j,m}x^k + \frac{\pi}{4} \mp \frac{\pi}{4k}\right)\right]$$

for $x>0, j=1, 2$.

For $x<0$ we have since $\rho'=\rho$, $\eta=0$,

$$u(x) = \frac{i\theta_2(\beta)}{c_{21}\theta_1(\beta) + c_{22}\theta_2(\beta)}\Omega(\beta,\xi).$$

Substituting in this the values of c_{21}, c_{22} as given by (10) we find that the fraction on the right reduces to $(-1)^j$ for $\rho=\rho_{j,m}$. Hence we have on comparing with (42)

$$u_{j,m}(x) = \mp u_{j,m}(\xi) \text{ for } x < 0, \qquad j = 1,2.$$

THEOREM 4. *If in equation* (1) *ν is an even integer and $\alpha=\beta$, the asymptotic forms of the characteristic functions are given for $x>0$ by* (44) *above. The functions $u_{1,m}(x)$ are odd functions and the functions $u_{2,m}(x)$ are even.*

Proceeding as in the previous case when $x=0$, we find in this case the values

$$(45)\qquad \begin{aligned} u_{1,m}(0) &= 0, & u'_{1,m}(0) &= C'_1 \rho_{1,m}^{1/2+1/(2k)}, \\ u_{2,m}(0) &= C_2 \rho_{2,m}^{1/2-1/(2k)}, & u'_{2,m}(0) &= 0. \end{aligned}$$

6. **The formal expansion of a function $f(x)$.** We shall suppose now, that the characteristic values are arranged and designated by a single subscript so that $|\rho_{m-1}| \leqq |\rho_m|$. Then two characteristic functions corresponding to distinct characteristic values are found in the usual way to satisfy the relation

$$\int_{-\alpha}^{\beta} x^\nu u_n(x) u_l(x)\,dx = 0, \qquad l \neq n.$$

It is clear that $\lambda=0$ is not a characteristic value, since by equation (1) this would imply $u''(x)\equiv 0$, and by (16) $u(x)\equiv 0$. Hence we have for any characteristic function u and the corresponding λ the relation

$$(46)\qquad x^\nu u^2 \equiv \frac{1}{2k\lambda}\frac{d}{dx}\{\lambda x^{\nu+1}u^2 + xu'^2 - uu'\},$$

which is an identity since $u(x)$ satisfies (1). Integrating it and applying (16) we obtain

$$\int_{-\alpha}^{\beta} x^\nu u_n^2(x)\,dx = \frac{1}{2k\lambda_n}[\beta u_n'^2(\beta) + \alpha u_n'^2(-\alpha)].$$

It is readily found by substituting on the right the asymptotic forms of the characteristic functions that this member and hence the integral on the left is different from zero for $n>N$, i.e. for $n>N$ the characteristic values are all simple. If this is true for all values of n, which is clearly the case when the characteristic functions are real, i.e. when ν is an integer, we may derive in the usual way the expansion associated with an arbitrary function $f(x)$, i.e.

$$\lim_{m\to\infty} \int_{-\alpha}^{\beta} t^\nu f(t)\phi_m(x,t)\,dt,$$

where

$$(47)\qquad \phi_m(x,t) = \sum_{n=1}^{m} \frac{u_n(x)u_n(t)}{\int_{-\alpha}^{\beta} t^\nu u_n^2(t)\,dt}.$$

Moreover if $G(x, t, \rho)$ is the Green's function of the system (1), (16), and γ is any semi-circle in the ρ-plane with center at $\rho=0$ which with the diameter

joining its end points encloses the characteristic values λ_n, $n=1, 2, \cdots, m$, and no others, then by familiar reasoning*

$$\phi_m(x,t) = \frac{k^2}{\pi i}\int_\gamma \rho G(x,t,\rho)d\rho. \tag{48}$$

7. **The function $G(x, t, \rho)$.** To avoid repetition in the discussion of the expansion problem we shall suppose that x, which we consider fixed, lies on the interval $0 \leqq x \leqq \beta$. Clearly the deductions and results for an $x<0$ will be entirely similar to those for $x>0$.

If $W[u_2, u_1]$ denotes the Wronskian of the functions u_2 and u_1, i.e.

$$W[u_2,u_1] = u_2u_1' - u_1u_2',$$

we find from formulas (3) and (3a)

$$W[u_2,u_1] = \frac{k\pi\rho^2}{2}x^k W[H^{(2)}_{1/(2k)}(\rho x^k), H^{(1)}_{1/(2k)}(\rho x^k)],$$

and it follows from the known formula for the Wronskian on the right [W, p. 76] that

$$W[u_2,u_1] \equiv 2k\rho i.$$

The classical formulas for $G(x, t, \rho)$† reduce therefore for the case in hand to

$$G(x,t,\rho) = \begin{cases} \dfrac{i\Omega(\beta,x)\Omega(t,-\alpha)}{2k\rho\Omega(\beta,-\alpha)} & \text{for } t<x, \\ \dfrac{i\Omega(\beta,t)\Omega(x,-\alpha)}{2k\rho\Omega(\beta,-\alpha)} & \text{for } t>x. \end{cases} \tag{49}$$

We observe now that because of the distribution of the characteristic values as shown in §4 it is possible to choose in the ρ-plane a sequence of circles Γ_m: $|\rho| = r_m$ having the following properties:

(i) as $m\to\infty$ over positive integral values, $r_m\to\infty$;

(ii) if γ_m denotes the semi-circle of Γ_m for which $-\delta \leqq \theta \leqq \pi-\delta$, then γ_m with the diameter joining its end points includes just m characteristic values;

(iii) there exists a constant $\chi>0$ such that every point of any arc γ_m is at a distance greater than χ from any characteristic value.

For the subsequent discussion we must derive the asymptotic forms of

* Birkhoff, these Transactions, vol. 9 (1908) p. 379. If some characteristic values are not simple we may consider $\phi_m(x, t)$ defined by (48) rather than by (47).

† Birkhoff, loc. cit., p. 378.

$G(x, t, \rho)$ on the semi-circles γ_m. With this in mind we proceed as follows. If ν is not an even integer let γ_m be divided into the arcs

$$\gamma_{m1}: -\delta \leqq \theta \leqq 0,$$
$$\gamma_{m2}: \quad 0 \leqq \theta \leqq \eta\pi,$$
$$\gamma_{m3}: \quad \eta\pi \leqq \theta \leqq \pi - \delta.$$

Consider first any portion of γ_{m2}, say γ'_{m2}: $0 \leqq \theta \leqq \theta' < \eta\pi$. On this arc $R(i\rho') > 0$ and hence we have

$$\Omega(\beta, -\alpha) \sim -i^{1+\eta} c_{11}\theta_2(\beta,\rho)\theta_1(\alpha,\rho')\left[1 - \frac{c_{21}\theta_1(\beta,\rho)}{c_{11}\theta_2(\beta,\rho)}\right].$$

Substituting from (4) in the bracket on the right we obtain for it the expression

$$\left[1 - \frac{c_{21}}{c_{11}} e^{2i\rho\beta^k}\left(1 + \frac{B(\rho)}{\rho}\right)\right], \tag{50}$$

and this is clearly asymptotic to 1 except in the neighborhood of $\theta = 0$. But it is known* that if $|\rho - \rho_m| > \chi$, where ρ_m is any root of the equation obtained by equating the expression (50) to zero, then the expression (50) itself is uniformly bounded from zero. Similarly if we consider γ''_{m2}: $0 < \theta' \leqq \theta \leqq \eta\pi$, then on this arc $R(i\rho) < 0$ and we have

$$\Omega(\beta, -\alpha) \sim i^{1+\eta} c_{11}\theta_1(\alpha,\rho')\theta_2(\beta,\rho)\left[1 + \frac{c_{12}\theta_2(\alpha,\rho')}{c_{11}\theta_1(\alpha,\rho')}\right].$$

The bracket in this expression is clearly asymptotic to 1 except in the neighborhood of $\theta = \eta\pi$ where it is uniformly bounded from zero since $|\rho - \rho_m| > \chi$. In the same way we find that $\Omega(\beta, -\alpha)$ is uniformly bounded from zero on the arcs γ_{m1} and γ_{m3}, while

$$\Omega(\beta, -\alpha) \sim \begin{cases} i^{1+\eta} c_{21}\theta_1(\beta,\rho)\theta_1(\alpha,\rho') & \text{for } -\delta \leqq \theta \leqq \theta'' < 0, \\ -i^{1+\eta} c_{12}\theta_2(\beta,\rho)\theta_2(\alpha,\rho') & \text{for } \eta\pi < \theta' \leqq \theta \leqq \pi - \delta. \end{cases}$$

If ν is an even integer, we have from (28)

$$\Omega(\beta, -\alpha) = -ic_{12}\theta_2(\beta)\theta_2(\alpha)\left[1 + \frac{c_{11}\theta_1(\alpha)}{c_{12}\theta_2(\alpha)} - \frac{c_{22}\theta_1(\beta)}{c_{12}\theta_2(\beta)} - \frac{\theta_1(\alpha)\theta_1(\beta)}{\theta_2(\alpha)\theta_2(\beta)}\right],$$

and the preceding argument may be applied to show that the bracket on

---

* Tamarkin, Rendiconti del Circolo Matematico di Palermo, vol. 34 (1912). Cf. also Wilder, loc. cit., p. 422.

the right is asymptotic to 1 for $0<\theta'\leqq\theta\leqq\pi-\delta$. An analogous formula is found for $\Omega(\beta, -\alpha)$ when $-\delta\leqq\theta\leqq\theta''<0$, while the quantity is uniformly bounded from zero on the semi-circles γ_m. It is clear from this that the subsequent discussion of the case ν an even integer may be included in the preceding general case by permitting η to take the value zero.

Substituting the asymptotic forms in the numerators of formulas (49) we find that when $0\leqq x\leqq\beta$, $G(x, t, \rho)$ is bounded uniformly for ρ on the arcs γ_m, and that the asymptotic forms of $G(x, t, \rho)$ may be obtained when $0\leqq x<\beta$ and $-\alpha<t<\beta$, from the following expressions:

when $-\delta\leqq\theta\leqq\theta'<0$,

$$(51)\qquad G(x,t,\rho)\sim\begin{cases}\dfrac{i\theta_2(x)\left[c_{21}\theta_1(t)-c_{11}\theta_2(t)\right]}{2k\rho c_{21}} & \text{for } t<x,\\[2ex] \dfrac{i\theta_2(t)\left[c_{21}\theta_1(x)-c_{11}\theta_2(x)\right]}{2k\rho c_{21}} & \text{for } t>x;\end{cases}$$

when $0<\theta'\leqq\theta\leqq\theta''<\eta\pi$,

$$(51a)\qquad G(x,t,\rho)\sim\begin{cases}\dfrac{i\theta_1(x)\left[c_{21}\theta_1(t)-c_{11}\theta_2(t)\right]}{2k\rho c_{11}} & \text{for } t<x,\\[2ex] \dfrac{i\theta_1(t)\left[c_{21}\theta_1(x)-c_{11}\theta_2(x)\right]}{2k\rho c_{11}} & \text{for } t>x;\end{cases}$$

and when $\eta\pi<\theta'\leqq\theta\leqq\pi-\delta$,

$$(51b)\qquad G(x,t,\rho)\sim\begin{cases}\dfrac{i\theta_1(x)\left[c_{22}\theta_1(t)-c_{12}\theta_2(t)\right]}{2k\rho c_{12}} & \text{for } t<x,\\[2ex] \dfrac{i\theta_1(t)\left[c_{22}\theta_1(x)-c_{12}\theta_2(x)\right]}{2k\rho c_{12}} & \text{for } t>x.\end{cases}$$

8. **A lemma.** For the discussion of integrals involving $G(x, t, \rho)$ we shall find the following lemma of use. We designate by $\gamma^{(1,2)}$ the portion of the semicircle γ_m for which $\theta'\leqq\theta\leqq\theta''$ and set $\rho=re^{i\theta}$.

LEMMA. *If $\psi(t)$ and $\Phi(t, \rho)$ are any integrable functions such that for $a\leqq t\leqq b$,*

(i) *$\psi(t)$ is real and $\psi(t)\geqq\psi>0$;* (ii) *$|\Phi(t, \rho)|<M$ for $\theta'\leqq\theta\leqq\theta''$, $r>N$,*

then the integral

$$I(t)=\int_{\gamma^{(1,2)}}\rho^{l}e^{\pm i\rho\psi(t)}\Phi(t,\rho)d\rho,$$

the ambiguous sign being chosen + when ρ is in the upper half plane, and − when ρ is in the lower half plane, satisfies the relation $|I(t)|\leqq M_1 r^l$.

Suppose for the moment that $\gamma^{(1,2)}$ is an arc of the quadrant $0 \leqq \theta \leqq \pi/2$. Then we have $\sin\theta \geqq \theta/2$ and hence

$$\left| e^{i\rho\psi(t)} \right| = e^{-r\psi(t)\sin\theta} \leqq e^{-r\psi(t)\theta/2},$$

and

$$\left| I(t) \right| \leqq Mr^{l} \int_{\theta'}^{\theta''} e^{-r\psi(t)\theta/2} r d\theta < \frac{2Mr^{l}}{\psi}. \tag{52}$$

If $\gamma^{(1,2)}$ is an arc of the quadrant $\pi/2 \leqq \theta \leqq \pi$ we have $\sin\theta \geqq (\pi-\theta)/2$ and the result follows in the same way, as it obviously does also if $-\pi/2 \leqq \theta \leqq 0$.

9. **The function $\phi_m(x,t)$ when $x \neq 0$.** With the use of the formulas above we shall show first that when x is considered as a parameter, $x > 0$,

$$\left| t^{l/4}\phi_m(x,t) \right| < M \tag{53}$$

for $-\alpha \leqq t \leqq x-\zeta$, or $x+\zeta \leqq t \leqq \beta$, where ζ is an arbitrary constant subject to $\zeta > 0$, $0 \leqq x-\zeta$, $x+\zeta \leqq \beta$.

Since $G(x,t,\rho)$ as a function of x or as a function of t satisfies the boundary conditions (16) we see from (48) that

$$\begin{aligned} \phi_m(-\alpha,t) &\equiv \phi_m(\beta,t) \\ &\equiv \phi_m(x,-\alpha) \equiv \phi_m(x,\beta) \equiv 0. \end{aligned}$$

We may restrict the further discussion therefore to the case $x < \beta$, $-\alpha < t < \beta$. Furthermore we may choose as the contours γ in (48) precisely the semicircles γ_m of §7. Consider first the case in which $-\alpha < t \leqq x-\zeta$. Then since by (4) and (9)

$$t^{l/4}\left[c_{21}\theta_1(t) - c_{11}\theta_2(t)\right] = e^{i\rho(x-\zeta)^k} B_1(x,t,\rho) \quad \text{when } \rho \text{ is on } \gamma_{m1},$$

and

$$t^{l/4}\left[c_{2,j-1}\theta_1(t) - c_{1,j-1}\theta_2(t)\right] = e^{-i\rho(x-\zeta)^k} B_j(x,t,\rho) \quad \text{when } \rho \text{ is on } \gamma_{mj},\ j = 2,3,$$

we obtain by substituting (51), (51a) and (51b) in (48)

$$\begin{aligned} t^{l/4}\phi_m(x,t) = \frac{k}{2\pi}\Bigg[& \frac{1}{c_{21}} \int_{\gamma_{m1}} e^{-i\rho[xk-(x-\zeta)^k]} B_1(x,t,\rho) d\rho \\ & + \sum_{j=2}^{3} \frac{1}{c_{1,j-1}} \int_{\gamma_{mj}} e^{i\rho[xk-(x-\zeta)^k]} B_j(x,t,\rho) d\rho \Bigg]. \end{aligned}$$

From this it follows by the lemma of §8 that (53) holds for $-\alpha < t \leqq x-\zeta$. A similar procedure establishes (53) when $x+\zeta \leqq t < \beta$.

Consider now the integral

$$I_m(x,t_1,t_2) = \int_{t_1}^{t_2} t^r \phi_m(x,t)dt, \quad \text{for } 0 < x < \beta. \tag{54}$$

We shall show that

$$|I_m(x,t_1,t_2)| < \frac{M}{r_m}, \quad \begin{array}{l}\text{uniformly for } -\alpha \leqq t_1 \leqq t_2 \leqq x-\zeta,\\ \text{or } x+\zeta \leqq t_1 \leqq t_2 \leqq \beta;\end{array} \tag{55}$$

$$I_m(x, -\alpha, x) = \tfrac{1}{2} + \epsilon(x,m)\ ; \tag{56}$$

$$I_m(x,x,\beta) = \tfrac{1}{2} + \epsilon(x,m)\ ; \tag{57}$$

$$|I_m(x,t_1,t_2)| < M \text{ uniformly for } -\alpha \leqq t_1 \leqq t_2 \leqq \beta, \tag{58}$$

where $\epsilon(x, m)$ is used as a generic symbol for a function which approaches zero as $m \to \infty$.

The function $G(x, t, \rho)$ as a function of t satisfies the equation (1). Hence, on substituting from (48) in (54) we may write the result

$$I_m(x,t_1,t_2) = -\frac{1}{\pi i}\int_{t_1}^{t_2}\int_{\gamma_m} G_{tt}(x,t,\rho)\frac{d\rho}{\rho}dt.$$

Interchanging the order of integration and integrating we obtain from this

$$I_m(x,t_1,t_2) = -\frac{1}{\pi i}\int_{\gamma_m}\{G_t(x,t_2,\rho) - G_t(x,t_1,\rho)\}\frac{d\rho}{\rho}, \tag{59}$$

$$\text{provided } -\alpha \leqq t_1 \leqq t_2 \leqq x, \quad \text{or } x \leqq t_1 \leqq t_2 \leqq \beta.$$

Now from formulas (8) and (13) we find that for $-\alpha \leqq t \leqq x-\zeta$

$$\frac{1}{\rho}[c_{21}u_1'(t) - c_{11}u_2'(t)] = e^{i\rho(x-\zeta)^k}B_1(x,t,\rho), \text{ for } \rho \text{ on } \gamma_{m1},$$

and

$$\frac{1}{\rho}[c_{2,j-1}u_1'(t) - c_{1,j-1}u_2'(t)] = e^{-i\rho(x-\zeta)^k}B_j(x,t,\rho), \text{ for } \rho \text{ on } \gamma_{mj},\ j = 2,3,$$

and hence from (51) and the lemma of §8

$$\begin{aligned}\int_{\gamma_m} G_t(x,t,\rho)\frac{d\rho}{\rho} &= \int_{\gamma_{m1}} e^{-i\rho[x^k-(x-\zeta)^k]}B_1(x,t,\rho)\frac{d\rho}{\rho}\\ &+ \sum_{j=2}^{3}\int_{\gamma_{mj}} e^{i\rho[x^k-(x-\zeta)^k]}B_j(x,t,\rho)\frac{d\rho}{\rho} = \frac{B(x,t,\rho)}{r_m}.\end{aligned} \tag{60}$$

Similar considerations lead to the same result when $x+\zeta \leqq t \leqq \beta$. This establishes the relation (55). If we set $\zeta = 0$ in (60) the exponential factor

reduces to unity. Obviously, however, the integrals are still bounded as $m\to\infty$. It follows that (58) is valid for $t\leqq x$, and a similar discussion establishes it for $t\geqq x$. To deduce (56) we observe that on substituting $t=x$ in the first of formulas (51), (51a), and (51b) and using the asymptotic forms (4) we obtain

$$(61)\qquad \int_{\gamma_m} G_t(x,x-,\rho)\frac{d\rho}{\rho} = \int_{\gamma_m}\left[-\frac{1}{2}+\frac{B(x,\rho)}{\rho}\right]\frac{d\rho}{\rho} = -\frac{\pi i}{2}+\epsilon(x,m).$$

Substituting this and the relation (60) for $t=-\alpha$ into (59) we obtain the result (56)

Since the Green's function has the property that

$$(62)\qquad G_t(x,x+,\rho) - G_t(x,x-,\rho) \equiv 1,$$

we obtain directly from (61)

$$\int_{\gamma_m} G_t(x,x+,\rho)\frac{d\rho}{\rho} = \frac{\pi i}{2}+\epsilon(x,m).$$

This together with (60) for $t=\beta$ yields (57).

10. **Convergence of the expansion at $x\neq 0$.** We prove the following theorem:

THEOREM 5. *If $f(x)$ is any function which is integrable in the sense of Lebesgue over the interval $-\alpha\leqq x\leqq\beta$, and if x is any inner point of this interval other than $x=0$, and if further the function $f(t)$ is of bounded variation in any neighborhood of the point $t=x$, then the expansion associated with $f(x)$ converges to*

$$\tfrac{1}{2}[f(x+)+f(x-)].$$

*At the end points $x=-\alpha$ or $x=\beta$ the expansion converges to zero.**

From (56) and (57) the expression $\frac{1}{2}[f(x+)+f(x-)]$ may be written in the form

$$(63)\qquad \lim_{m\to\infty}[f(x+)I_m(x,x,\beta)+f(x-)I_m(x,-\alpha,x)].$$

On the other hand the sum of the first m terms of the expansion associated with $f(x)$ is by (47)

$$\int_{-\alpha}^{\beta} t^r f(t)\phi_m(x,t)dt.$$

It remains to be shown that under the existing hypotheses the difference of these expressions converges to zero. This will follow if

* The convergence at $x=0$ is covered by a later theorem.

2,

SUGGESTIONS TO AUTHORS

Much needless expense and many errors can be avoided. The editors of several mathematical journals have agreed upon the following suggestions.

1. Typewrite words and the very simplest formulas only.
2. *Do not* try to typewrite any complex formulas. Write them.
3. Keep a copy, and send the editors two copies, if you can.
4. *Do not* underline any symbols or any formulas.
5. Underline theorems with blue pencil (avoid ink).
6. Follow our recent styles in abbreviations, footnotes, etc.
7. Write carefully the (often misunderstood) capitals C K P S V W X Z.
8. Write ϵ, not ε. Write very carefully $\gamma\ \eta\ \kappa\ \lambda\ \nu\ \tau\ \upsilon\ \chi\ \omega$.
9. Among Greek capitals, use only $\Gamma\ \Delta\ \Theta\ \Lambda\ \Xi\ \Pi\ \Sigma\ \Phi\ \Psi\ \Omega$.
10. Punctuate carefully, especially in formulas; thus: $1, 2, \cdots, n$.
11. Use the solidus (/) to avoid fractions in solid lines.
12. Use fractional exponents to avoid root signs everywhere.
13. Use extra symbols to avoid complicated exponents.
14. In typewritten formulas, 1 means "one"; to indicate "ell" in formulas, backspace and overprint /; thus: $1/$. Similarly, 0 means "zero"; to indicate "cap O," backspace and overprint period; thus: $0.$
15. Avoid a dash over a letter, except for those shown below.
16. Some samples of unusual types available on monotype machines follow. A more complete list of all such types will be sent on request.

Light Face Greek—$\alpha\ \beta\ \gamma \cdots$ (all) A B $\Gamma \cdots$ (all).

★ Light Greek Superiors—$^{\alpha}$ and $^{\alpha\ \beta\ \gamma \cdots}$ (all except ^{o}).

★ Light Greek Inferiors—$_{\alpha\ \beta\ \gamma \cdots}$ and $_{a\ e\ o\ x}$ (all except $_{o}$).

• Boldface Greek—$\boldsymbol{\alpha\ \beta\ \gamma\ \delta\ \varepsilon\ \zeta\ \eta\ \theta\ \kappa\ \lambda\ \mu\ \nu\ \xi\ \pi\ \varrho\ \sigma\ \omega}$ and $\boldsymbol{\Omega}$.

• Lightface German—$\mathfrak{a\ b\ c\ d\ e\ f\ p\ q}$ $\mathfrak{A\ B\ C\ D\ E\ F\ G\ H\ J\ K\ L\ M\ P\ Q\ R}$ $\mathfrak{S\ T\ U\ V\ W\ X\ Y\ Z}$.

• Boldface German—$\mathfrak{b}$ $\mathfrak{A\ B\ D}$.

Script (special font) $\mathcal{A\ B\ C} \cdots$ (all). No lower case manufactured.

• Hebrew—ℵ ℶ ℷ ℸ troublesome to handle.

★ Dashed Italics—$\overline{A}\ \overline{a}\ \overline{B}\ \overline{b}\ \overline{C}\ \overline{c}\ \overline{E}\ \overline{e}\ \overline{F}\ \overline{f}\ \overline{G}\ \overline{g}\ \overline{H}\ \overline{I}\ \overline{i}\ \overline{j}\ \overline{K}\ \overline{k}\ \overline{M}\ \overline{m}\ \overline{N}\ \overline{n}\ \overline{o}\ \overline{P}\ \overline{p}$ $\overline{q}\ \overline{R}\ \overline{r}\ \overline{s}\ \overline{t}\ \overline{u}\ \overline{v}\ \overline{X}\ \overline{x}\ \overline{Y}\ \overline{y}\ \overline{Z}\ \overline{z}$.

★ Tilda Italics—$\tilde{A}\ \tilde{a}\ \tilde{e}\ \tilde{N}\ \tilde{n}\ \tilde{O}\ \tilde{o}\ \tilde{u}\ \tilde{y}$.

★ Tilda Greek—$\tilde{\alpha}\ \tilde{\varepsilon}\ \tilde{\eta}\ \tilde{\omega}\ \tilde{\omega}$.

★ Dashed Greek—$\bar{\alpha}\ \bar{\beta}\ \bar{\gamma}\ \bar{\delta}\ \bar{\eta}\ \bar{\theta}\ \bar{\mu}\ \bar{\nu}\ \bar{\rho}\ \bar{\omega}\ \bar{\Gamma}$.

★ Dotted Italic—$\dot{a}\ \dot{a}\ \dot{c}\ \dot{e}\ \dot{g}\ \dot{i}\ \dot{m}\ \dot{n}\ \dot{q}\ \dot{r}\ \dot{s}\ \dot{u}\ \dot{x}\ \dot{x}\ \dot{y}\ \dot{y}\ \dot{z}\ \dot{z}$.

★ Dotted Greek—$\dot{\eta}\ \dot{\eta}\ \dot{\theta}\ \dot{\theta}\ \dot{\xi}\ \dot{\psi}\ \dot{\psi}\ \ddot{\omega}$ (single dotted $\zeta\ \phi\ \delta\ \beta\ \gamma$; double dotted γ readily available).

• Additional characters readily available at small cost.

★ Matrices for additional characters are made upon special orders and necessitate a delay of from four to eight weeks and average expense of $4.50 per matrix.

re
m
li
t
(

$$\lim_{m\to\infty} D_j(x,m) = 0, \qquad j = 1,2, \tag{64}$$

where

$$D_1(x,m) = \int_{-\alpha}^{x} t^p[f(t) - f(x-)]\phi_m(x,t)dt,$$
$$D_2(x,m) = \int_{x}^{\beta} t^p[f(t) - f(x+)]\phi_m(x,t)dt. \tag{65}$$

Consider the first of these expressions (65). By hypothesis there exists an interval $(x-\zeta, x)$ in which $f(t)$ is of bounded variation. Hence we may write in this interval

$$f(t) - f(x-) = \chi_1(t) - \chi_2(t),$$

where $\chi_j(t)$, $j=1, 2$, are bounded, positive and decreasing functions and $\chi_j(x-)=0$. Hence if ϵ is arbitrary, $\epsilon>0$, there exists a positive constant $\zeta_1<\zeta$ such that

$$0 \leqq \chi_j(t) < \epsilon \text{ for } x - \zeta_1 \leqq t \leqq x, \qquad j = 1,2. \tag{66}$$

We have now

$$D_1(x,m) = \int_{-\alpha}^{x-\zeta_1} t^p[f(t) - f(x-)]\phi_m(x,t)dt - \sum_{j=1}^{2}(-1)^j \int_{x-\zeta_1}^{x} t^p\chi_j(t)\phi_m(x,t)dt. \tag{67}$$

By virtue of (53), (55) and Lebesgue's* lemma the first of these integrals converges to zero as $m\to\infty$. By the second law of the mean, however, the remaining integrals in (67) can be written in the form

$$\chi_j(x-\zeta_1)\int_{x-\zeta_1}^{t_j} t^p\phi_m(x,t)dt, \text{ where } x - \zeta_1 < t_j < x, \qquad j = 1,2,$$

and by (58) and (66) this is less than $M\epsilon$. Hence for m sufficiently large

$$|D_1(x,m)| < 3M\epsilon$$

and the first of relations (64) is established. The discussion of the expression $D_2(x, m)$ is precisely similar and hence we may consider the theorem established.

11. **The function $\phi_m(x, t)$ when $x=0$.** The discussion above was based on the assumption $x\neq 0$. We consider now the excepted case. The function $\phi_m(x, t)$ is readily found to be unbounded as a function of m when $x=0$. However, we shall show that when $x=0$ the integral (54) satisfies the relations

* Lebesgue, Annales de la Faculté des Sciences de Toulouse, (3), vol. 1 (1909), p. 52.

$$(68)\quad |I_m(0,t_1,t_2)| < \frac{M}{r_m^{1/2+1/(2k)}}, \quad \text{uniformly for } -\alpha \leqq t_1 \leqq t_2 \leqq -\zeta \text{ or } \zeta \leqq t_1 \leqq t_2 \leqq \beta;$$

$$(69)\quad I_m(0,-\alpha,0) = \frac{1}{2} + i\sigma + \epsilon(m), \text{ where } \sigma = \frac{1}{2}\left(\eta \tan\frac{\eta'\pi}{2k} - \eta' \tan\frac{\eta\pi}{2k}\right);$$

$$(70)\quad I_m(0,0,\beta) = \frac{1}{2} - i\sigma + \epsilon(m);$$

$$(71)\quad |I_m(0,t_1,t_2)| < M \quad \text{uniformly for } -\alpha \leqq t_1 \leqq t_2 \leqq \beta.$$

We observe that when ν is an integer, $\sigma=0$.

The formula (59) is valid for $x=0$. Hence we consider the integral (60) for $x=0$. We have from (51), (51a) and (51b) for $-\alpha \leqq t < 0$

$$(72)\quad \int_{\gamma_m} G_t(0,t,\rho)\frac{d\rho}{\rho} = \frac{-i^\nu}{2k}\left[\frac{1}{c_{21}}\int_{\gamma_{m1}} \frac{u_2(0)u_2'(\tau,\rho')}{\rho}\frac{d\rho}{\rho} + \sum_{j=2}^{3}\frac{(-1)^j}{c_{1,j-1}}\int_{\gamma_{mj}}\frac{u_1(0)u_{4-j}'(\tau,\rho)}{\rho}\frac{d\rho}{\rho}\right]$$

where $\tau = te^{-\pi i}$. If $t \leqq -\zeta$ we may substitute the asymptotic forms of $u_j'(\tau,\rho')$. This gives to the right hand member the form

$$\int_{\gamma_{m1}} e^{-i\rho'\tau k}\frac{B_1(\tau,\rho)}{\rho^{1/2+1/(2k)}}d\rho + \int_{\gamma_{m2}} e^{-i\rho'\tau k}\frac{B_2(\tau,\rho)}{\rho^{1/2+1/(2k)}}d\rho + \int_{\gamma_{m3}} e^{i\rho'\tau k}\frac{B_3(\tau,\rho)}{\rho^{1/2+1/(2k)}}d\rho$$

to which we may apply the lemma of §8. From this together with the discussion for $\zeta \leqq t \leqq \beta$, formula (68) follows. On the other hand if $-\zeta \leqq t \leqq 0$, we may turn to formula (12) for the form of $u_j'(\tau,\rho')$. The right hand member of (72) is in this case of the form

$$\sum_{j=1}^{3}\int_{\gamma_{mj}} B_j(\tau,\rho)\frac{d\rho}{\rho},$$

and relation (71) follows.

For the determination of (69) we consider the integral (72) for $t=0-$. Substituting the values given by (15) in (69) and (70) and employing (14) we obtain

$$\int_{\gamma_m} G_t(0,0-,\rho)\frac{d\rho}{\rho} = \frac{-1}{2\sin\dfrac{\pi}{2k}}\left\{\left[\frac{c_{11}}{c_{21}} + ie^{-\pi i/(2k)}\right]\int_{\gamma_{m1}}\frac{d\rho}{\rho} + \left[\frac{c_{21}}{c_{11}} - ie^{\pi i/(2k)}\right]\int_{\gamma_{m2}}\frac{d\rho}{\rho} + \left[\frac{c_{22}}{c_{21}} - ie^{\pi i/(2k)}\right]\int_{\gamma_{m3}}\frac{d\rho}{\rho}\right\}.$$

But

$$\int_{\gamma_{m1}} \frac{d\rho}{\rho} = \delta i, \quad \int_{\gamma_{m2}} \frac{d\rho}{\rho} = \eta\pi i, \text{ and } \int_{\gamma_{m3}} \frac{d\rho}{\rho} = (\eta'\pi - \delta)i,$$

while from (10)

$$\frac{c_{11}}{c_{21}} = -i\left(\cos\frac{\pi}{2k} + \tan\frac{\eta\pi}{2k}\sin\frac{\pi}{2k}\right),$$

$$\frac{c_{21}}{c_{11}} = i\left(\cos\frac{\pi}{2k} + \tan\frac{\eta'\pi}{2k}\sin\frac{\pi}{2k}\right),$$

$$\frac{c_{22}}{c_{21}} = i\left(\cos\frac{\pi}{2k} - \tan\frac{\eta\pi}{2k}\sin\frac{\pi}{2k}\right).$$

With these values we obtain

$$\int_{\gamma_m} G_t(0,0-,\rho)\frac{d\rho}{\rho} = -\pi i\left\{\frac{1}{2} + \frac{i}{2}\left(\eta\tan\frac{\eta'\pi}{2k} - \eta'\tan\frac{\eta\pi}{2k}\right)\right\},$$

and this leads directly to (69). The formula (70) follows from (69) because of (62) as in §9.

12. **The expansion at $x=0$.** We prove the following theorem:

THEOREM 6. *If $f(x)$ is any function which is of bounded variation on the interval $-\alpha \leqq x \leqq \beta$ then the expansion associated with $f(x)$ converges when $x=0$ to*

$$[\tfrac{1}{2} - i\sigma]f(0+) + [\tfrac{1}{2} + i\sigma]f(0-), \tag{73}$$

where σ depends on ν as given in (69), *and vanishes when ν is an integer.*

Proceeding as in §9, we write (73) in the form (63) with $x=0$, and construct the quantities $D_j(0, m)$ given by (65). We must show that (64) follows when $x=0$. Consider $D_1(0, m)$. Since $f(t)-f(0-)$ is of bounded variation we may write it in the form

$$\chi_1(t) - \chi_2(t),$$

where $\chi_j(t)$ are positive and decreasing, and

$$|\chi_j(t)| < M. \tag{74}$$

Moreover since $\chi_j(0-)=0$, we may choose ϵ arbitrarily and determine ζ so that

$$|\chi_j(t)| < \epsilon, \text{ when } -\zeta \leqq t \leqq 0. \tag{75}$$

Then

$$D_1(0,m) = -\sum_{j=1}^{2}(-1)^j \int_{-\alpha}^{-\zeta} t^\nu \chi_j(t)\phi_m(0,t)dt - \sum_{j=1}^{2}(-1)^j \int_{-\zeta}^{0} t^\nu \chi_j(t)\phi_m(0,t)dt.$$

Applying the second law of the mean to each of these integrals, we obtain the relation

$$D_1(0,m) = -\sum_{j=1}^{2}(-1)^j \chi_j(-\alpha) \int_{-\alpha}^{T_j} t^\nu \phi_m(0,t)dt - \sum_{j=1}^{2}(-1)^j \chi_j(-\zeta) \int_{-\zeta}^{t_j} t^\nu \phi_m(0,t)dt,$$

where $-\alpha < T_j < -\zeta$, $-\zeta < t_j < 0$. By (74), (68), (75), and (71), however, this yields

$$|D_1(0,m)| \leqq \frac{M}{r_m^{1/2+1/(2k)}} + 2M\epsilon < 3M\epsilon \text{ for } m \geqq m_1.$$

This establishes the fact that

$$\lim_{m\to\infty} D_j(0,m) = 0,$$

for $j=1$, and a similar argument establishes it for $j=2$. Hence Theorem 6 is proved.

13. **The case $\alpha=0$.** We consider now briefly the case in which $x=0$ is an end point of the interval. The boundary conditions are then

$$u(\beta) = 0, \quad u(0) = 0,$$

and the characteristic equation is

$$\Omega(\beta,0) = 0. \tag{76}$$

From (4) and (15) the equation (76) is of the form

$$\rho^{1/2+1/(2k)}[\theta_1(\beta) + ie^{\pi i/(2k)}\theta_2(\beta)] = 0. \tag{77}$$

The observation of §6 that $\rho=0$ is not a characteristic value is valid for the case in hand, and since the remaining factor in (77) is precisely the first factor of (31) the characteristic values ρ_m (or λ_m) are given, regardless of the character of ν by precisely the formula (32) (or (33)) with $j=1$. It follows that the characteristic functions also are given for any value of ν by the formulas for $u_{1,m}(x)$ in (43), (44), and (45).

Since $x=0$ is now an end point of the interval the expansion converges at this point to zero. For $0<x\leqq\beta$ the discussion is precisely that of §§6, 7, 9 and 10 with α replaced by 0, and the expansion theorem is accordingly the theorem as stated in §10.

UNIVERSITY OF WISCONSIN,
MADISON, WIS.

TRIPLE AND MULTIPLE SYSTEMS, THEIR GEOMETRIC CONFIGURATIONS AND GROUPS*

BY
ARNOLD EMCH

I. INTRODUCTION

The fact that every finite group may be represented as a substitution group of a certain degree m makes it possible to construct in every case certain configurations in a projective space of $m-1$ or lower dimensions whose points correspond uniquely to the substitutions of the group. Such configurations are representatives of the groups and are characterised by geometric properties which reflect the properties of the group. Such representations in certain particular cases are numerous and have been known for a long time. From this standpoint I have considered the symmetric group of degree m on a hyperquadric in S_{m-1}.† It is thus obvious that any finite group may be represented by a certain configuration (with characteristic geometric properties) on a hyperquadric. Whether such a representation would in every case be mathematically profitable is an open question. While there is thus indicated a general method of approach to the geometrical problem of finite group theory, there may be developed special methods of representation of finite groups in connection with particular problems connected with group theory, which seem to be of sufficient interest and importance to deserve attention. This is the case with the theory of triple and multiple systems, where geometric considerations are found in the literature only in a few instances.‡ It was Kirkman§ in 1844, and Steiner in 1853 who first stated the problem of triple systems and certain multiple systems in its most elementary aspect as it was revealed by the study of the configuration of

---

* Presented to the Society, September 8, 1926, and April 6, 1928; received by the editors March 31, 1928.

† *Some geometric applications of symmetric substitution groups*, American Journal of Mathematics, vol. 45 (1923), pp. 192–207.

‡ L. Heffter, *Ueber das Problem der Nachbargebiete*, Mathematische Annalen, vol. 38 (1891), pp. 477–508.

De Vries, *Zur Theorie der Tripel Systeme*, Rendiconti del Circolo Matematico di Palermo, vol. 8 (1894), pp. 222–226. *Ueber polyedrale Configurationen*, Mathematische Annalen, vol. 34 (1889) pp. 237–246.

§ *On a problem in combinations*, The Cambridge and Dublin Mathematical Journal, vol. 2 (1847), pp. 191–204.

the 28 double tangents of the quartic.* A detailed account of this connection is given by Noether in an important paper† which appeared in 1879. Netto pays attention to the algebraic properties of triple systems and the closely associated triple equations and their groups. Since that time triple systems have been studied by a number of investigators mostly with reference to existence proofs, formation and group properties of such systems.‡

That a vast number of similar systems, for example triple systems with double couples, quadruple systems containing every triple once, in general n-tuple systems of various types, may be formed becomes evident without much effort. There seems to be little that has been done along this line of generalization. Inasmuch as with every well defined system of this sort there is associated a finite group which dominates it and as the number of such systems is unlimited, there exists a large class of finite groups which bear conversely the characteristics of these systems. Each triple or multiple system and couples of such may be represented by a certain spatial configuration to which is attached a related finite group. It is the purpose of this paper to show by a few simple examples how such configurations and their groups may be constructed.

II. Triple systems

1. **Simple triple systems.** As systems of this sort, their description and some of their properties, at least for lower orders, are fairly well known I shall not enter into a detailed discussion of them, and shall present merely a few additional circumstances which seem to be worthy of notice.

It is well known that triple systems Δ_m^2 on m elements exist only, and in every case, for numbers of the form $m=6m+1$, and $m=6m+3$. Instead

---

* *Combinatorische Aufgabe*, Journal für reine und angewandte Mathematik, vol. 45 (1853), pp. 181–182.

† *Ueber die Gleichungen achten Grades und ihr Auftreten in der Theorie der Curven vierter Ordnung*, Mathematische Annalen, vol. 15 (1879), pp. 89–110.

‡ E. Netto, *Zur Theorie der Tripelsysteme*, Mathematische Annalen, vol. 42 (1893), pp. 141–152.

E. H. Moore, *Concerning triple systems*, Mathematische Annalen, vol. 43 (1893), pp. 271–285.

F. N. Cole, *The triad systems of thirteen letters*, these Transactions, vol. 14 (1913), pp. 1–5.

H. S. White, *Triple systems as transformations, and their paths among triads*, these Transactions, vol. 14 (1913), pp. 6–13.

Louise D. Cummings, *On a method of comparison for triple systems*, these Transactions, vol. 15 (1914), pp. 311–327.

H. S. White, *The multitude of triad systems on 31 letters*, these Transactions, vol. 16 (1915), pp. 13–19.

S. Bays, *Sur les systèmes cycliques de Steiner*, Annales Scientifiques de l'Ecole Normale Supérieure, (3), vol. 4 (1923), pp. 55–96.

In these papers may also be found additional references.

of using the medium of hyperquadric or other forms of possible graphic representation we may make use of the simple device of representing the m elements by m generic points in ordinary space S_3. *Then to every triple system corresponds a certain open polyhedral configuration of triangles on these m points. To two triple systems Δ_m^3 without a common triple corresponds a closed polyhedral configuration, whose mapping upon simple closed surfaces will be explained subsequently in case of $m=7$ and $m=9$.*

Δ_7^3. For certain purposes (determination of the order of the group, etc.) it is desirable, although not necessary, to have some sort of a generalized matrix from which we may write down a triple system in a very simple and effective manner. The terms for $m=7$, for seven elements 1, 2, $\cdots$, 7, may be written in any order $a, b, \cdots, g$ in the form

$$\left\| \begin{matrix} a & b & c \\ & d & \\ e & f & g \end{matrix} \right\|$$

from which the 7 triples are immediately obtained as (1) the first line, (2) the second column, (3) and (4), the two diagonals, (5) and (6), the first and third column each together with f, (7) the triangle $b\ e\ g$; i.e., $a\ b\ c$, $b\ d\ f$, $a\ d\ g$, $c\ d\ e$, $a\ e\ f$, $c\ g\ f$, $b\ e\ g$. Each such form or "matrix" may be identified by a substitution

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ a & b & c & d & e & f & g \end{pmatrix}.$$

It is easy to write down the 24 "matrices" with the central letter d unchanged which produce the same triple system. The 24 corresponding substitutions form a subgroup of order 24, and as the central d may be replaced by any of the remaining letters we get a group of $24 \cdot 7 = 168$ substitutions, the simple G_{168} of this order. As the structure of this group is well known, details of its construction by the foregoing scheme are not necessary.

There are altogether $7!:168=30$ distinct Δ_7^3's and any of the 35 triples is always common to 6 distinct Δ_7^3's. But it is not possible to form 5 distinct Δ_7^3's which together contain all of the 35 triples, as has already been established by Cayley.*

Δ_9^3. For the generating "matrix" we may use the square arrangement

$$\left\| \begin{matrix} a & b & c \\ d & e & f \\ g & h & i \end{matrix} \right\|.$$

* Papers, vol. 1, pp. 481–484.

The 12 triples obtained from the three rows, the three columns and the 6 terms of the ordinary determinant expansion form a triple system $a\,b\,c,\ d\,e\,f,\ g\,h\,i,\ a\,d\,g,\ b\,e\,h,\ c\,f\,i,\ a\,e\,i,\ a\,f\,h,\ b\,d\,i,\ c\,d\,h,\ b\,f\,g,\ c\,e\,g$, on 9 letters or numbers, in which $a, b, \cdots, i$ denote the 9 numbers $1, 2, \cdots, 9$ in any order. Thus the particular arrangement of the "matrix" is equivalent to the substitution

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ a & b & c & d & e & f & g & h & i \end{pmatrix}.$$

To determine the group which leaves the triple system invariant we notice first that the Δ_9^3 as a whole does not change when any two rows, or any two columns are interchanged. This gives a subgroup of 36 substitutions which leave Δ_9^3 invariant.

But there are 11 other distinct matrices and distinct from any attached to the G_{36}, which also leave Δ_9^3 invariant, namely (including the first)

$$\left\|\begin{matrix} a & b & c \\ d & e & f \\ g & h & i \end{matrix}\right\| \quad \left\|\begin{matrix} a & d & g \\ b & e & h \\ c & f & i \end{matrix}\right\| \quad \left\|\begin{matrix} a & e & i \\ b & f & g \\ c & d & h \end{matrix}\right\| \quad \left\|\begin{matrix} a & f & h \\ b & d & i \\ c & e & g \end{matrix}\right\|$$

$$\left\|\begin{matrix} a & b & c \\ e & f & d \\ i & g & h \end{matrix}\right\| \quad \left\|\begin{matrix} a & d & g \\ e & h & b \\ i & c & f \end{matrix}\right\| \quad \left\|\begin{matrix} a & e & i \\ f & g & b \\ h & c & d \end{matrix}\right\| \quad \left\|\begin{matrix} a & f & h \\ d & i & b \\ g & c & e \end{matrix}\right\|$$

$$\left\|\begin{matrix} a & b & c \\ f & d & e \\ h & i & g \end{matrix}\right\| \quad \left\|\begin{matrix} a & d & g \\ h & b & e \\ f & i & c \end{matrix}\right\| \quad \left\|\begin{matrix} a & e & i \\ g & b & f \\ d & h & c \end{matrix}\right\| \quad \left\|\begin{matrix} a & f & h \\ i & b & d \\ e & g & c \end{matrix}\right\|.$$

With each of these are connected as above 36 substitutions which leave Δ_9^3 invariant. These together form all the substitutions which leave the given Δ_9^3 invariant, and they consequently form a G_{432} of order $36 \cdot 12 = 432$, which is known as the group of the triple equation of degree 9. There are, of course, several ways of generating this group. For instance if we set

$$S_1 = 123 \cdot 456 \cdot 789,$$
$$S_2 = 147 \cdot 258 \cdot 369,$$
$$S_3 = 159 \cdot 267 \cdot 348,$$
$$S_4 = 168 \cdot 249 \cdot 357,$$

then $G_9(S_1, S_2, S_3, S_4)$ is a non-cyclic abelian group of order 9 and the G_{432} is the holomorph of G_9.

Each triple is common to 120 distinct triple systems.

If we consider for example the triple 123 fixed, then the remaining numbers form 6! permutations $a\ b\ c\ d\ e\ f$, each giving the second and third row in the "matrix," namely $a\ b\ c$ and $d\ e\ f$. But as these may be interchanged and a cycle of order 3 applied to each without changing Δ_9^3, there are evidently only $6!:6=120$ permutations which with 123 give distinct Δ_9^3's. Now it is clear that two triples with a common couple, for example 123 and 124, cannot belong to the same Δ_9^3. Moreover every couple must occur in every Δ_9^3 just once. Hence all Δ_9^3's can be classified in various ways into 7 classes of 120 triple systems, each class being characterised by a common triple. Thus 123, 124, 125, 126, 127, 128, 129 are each common to 120 Δ_9^3's. Thus *there are altogether* $7\cdot 120=840$ *distinct triple systems on* 9 *elements*. This checks up with the fact that $9!:432=840$.

Now in case of 9 the remarkable circumstance presents itself that *all* 84 *triples may be distributed completely and each only once in* 7 *triple systems. This may be called a complete set of triple systems.*

That this is not always possible is clear from the case of $m=7$. Such a complete set is for example obtained from the 7 matrices

$$\left\|\begin{matrix}1&2&3\\4&5&6\\7&8&9\end{matrix}\right\|\quad\left\|\begin{matrix}1&2&4\\3&5&9\\7&6&8\end{matrix}\right\|\quad\left\|\begin{matrix}1&2&5\\3&7&8\\9&4&6\end{matrix}\right\|\quad\left\|\begin{matrix}1&2&6\\3&4&7\\8&5&9\end{matrix}\right\|\quad\left\|\begin{matrix}1&2&7\\3&6&8\\4&9&5\end{matrix}\right\|\quad\left\|\begin{matrix}1&2&8\\4&3&9\\5&7&6\end{matrix}\right\|\quad\left\|\begin{matrix}1&2&9\\7&3&6\\8&4&5\end{matrix}\right\|.$$

As there are 840 distinct Δ_9^3's, and from the fact that there exists at least one complete set, it is easily seen that a great number of such complete sets may be formed, and that each Δ_9^3 may belong to more than one complete set. So far I have not been able to establish necessary and sufficient conditions on the number m for which complete sets of triple systems are possible. The original complete set of Δ_9^3's were found by trial.

2. **Triple systems with double couples.** n elements form $\frac{1}{6}n(n-1)(n-2)$ triples and $\frac{1}{2}n(n-1)$ couples. Let Δ be a triple system containing every couple twice and let x be the number of triples of the system. As every triple contains 3 couples,

$$3x=\frac{2\cdot n(n-1)}{2}=n(n-1),$$

and $x=\frac{1}{3}n(n-1)$. Hence n must have either the form $n=3m$, or $n=3m+1$, and

$$x=m(3m-1),\quad\text{or}\quad x=(3m+1)m.$$

For $m=1$ we get in the first case $m=3$, $x=2$, a triangle 123 counted twice; in the second case $m=4$, $x=4$, or the tetrahedron 123, 124, 134, 234.

The groups which leave this dihedron and tetrahedron invariant are of course the dihedral group on three letters, or the symmetric group, and the tetrahedral group.

For $m=2$, we have for two cases $n=6$, $x=10$ and $n=7$, $x=14$.

In general, for $m=2\mu$,

$$n = 6\mu, \qquad x = 2\mu(6\mu - 1),$$

or

$$n = 6\mu + 1, \quad x = 2\mu(6\mu + 1).$$

For $m=2\mu+1$,

$$n = 6\mu + 3, \quad x = 2(2\mu + 1)(3\mu + 1),$$

or

$$n = 6(\mu + 1), \quad x = 2(2\mu + 1)(3\mu + 2).$$

In case of an even n, i.e., $n=6\mu$, or $n=6(\mu+1)$ we have triple systems with double couples which form closed surfaces. In case of odd numbers, $n=6\mu+1$, $n=6\mu+3$, we have the same number as in triple systems with single couples, but twice the number of triples. Hence if we take two such triple systems with no common triple we have a triple system with double couples.

Thus when $\mu=1$, $n=7$, $x=14$. When $\mu=1$, in the other case, $n=6+3=9$, $x=24$.

3. **Triple system of order 7 with double couples.** Such a system may be constructed by taking two ordinary triple systems of order 7 with single couples without a common triple. For example

$$\left\| \begin{matrix} 1 & 2 & 3 \\ & 4 & \\ 5 & 6 & 7 \end{matrix} \right\| \equiv \begin{matrix} 1\,1\,1\,2\,2\,3\,3 \\ 2\,4\,5\,4\,5\,4\,6 \\ 3\,7\,6\,6\,7\,5\,7, \end{matrix}$$

$$\left\| \begin{matrix} 2 & 1 & 6 \\ & 4 & \\ 7 & 3 & 5 \end{matrix} \right\| \equiv \begin{matrix} 1\,1\,1\,2\,2\,3\,4 \\ 2\,3\,5\,3\,4\,5\,6 \\ 6\,4\,7\,7\,5\,6\,7. \end{matrix}$$

Each system is invariant in a simple G_{168}. There are 7 substitutions which permute the two systems, namely

$$S_1 = (27)(35)(46) \text{ leaves } (1),$$
$$S_2 = (15)(34)(67) \quad " \quad (2),$$
$$S_3 = (16)(25)(47) \quad " \quad (3),$$
$$S_4 = (12)(36)(57) \quad " \quad (4),$$
$$S_5 = (14)(26)(37) \quad " \quad (5),$$
$$S_6 = (17)(45)(23) \quad " \quad (6),$$
$$S_7 = (13)(24)(56) \quad " \quad (7)$$

in each system invariant. Then the product of each of these is a cyclic substitution of order 7 and all these products are the powers of one of these. For example

$$S_1S_2 = (1\ 5\ 4\ 7\ 2\ 6\ 3) = T,$$
$$S_3S_5 = (1\ 2\ 5\ 6\ 4\ 3\ 7) = T^4.$$

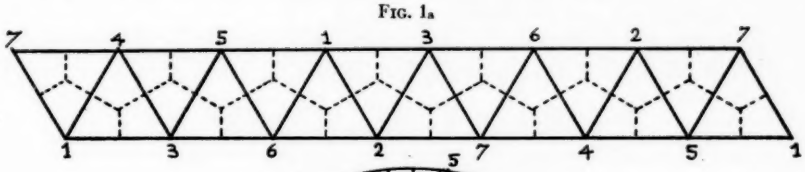
T and its powers transform the two triple systems into themselves. Hence this is a subgroup

$$G_{14} = 1, S_1, \cdots, S_7, T, T^2, \cdots, T^6$$

which leaves the triple system Δ_7^3 with double couples invariant.

If we let 1, 2, $\cdots$, 7 represent generic points in a space of three dimensions, so that each triple is represented by a triangle, then the two systems form a closed surface such that at each of the 21 couples there are two adjacent triangles. This surface can be developed in the manner indicated by the Fig. 1a and then, by uniting the edges $\overline{17}$, Fig. 1b, formed into a cyclinder. By deformation the upper rim can be shifted in a clockwise direction, so that corresponding equal numbers are in generatrices of the deformed cylinder.

FIG. 1a



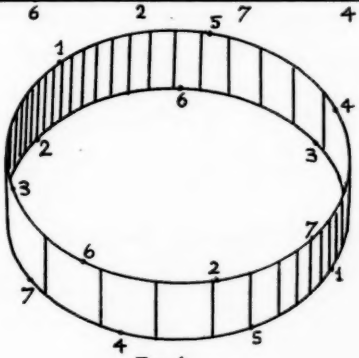


FIG. 1b

Then the cylinder can be deformed into a torus, by uniting equal numbers. By joining the centers of the triangles as indicated in the figure we obtain 7 hexagons on the torus with the seven points as centers. For example, the centers of the triangles 134, 147, 175, 156, 162, 123 form a hexagon with

center 1. The substitution S_1 leaves 1 invariant and permutes the sides of the hexagon cyclically: 156, 162, 123, 134, 147, 175. The substitution T changes the hexagon into the hexagon with center (5): 517, 572, 524, 543, 536, 561. The mapping process of the polyhedron of 14 triangles upon a torus is also verified by the formula

$$k = V + F - S - 7 + 14 - 21 = 0.$$

Moreover $2-k=2=2p$, $p=1$.

The hexagon with 1 as a center has the cycle $U=(3\,4\,7\,5\,6\,2)$ of order 6, which interchanges the sets of triple systems as follows:

$$\begin{array}{ccc} 1\,1\,1\,3\,3\,4\,4 & & 1\,1\,1\,3\,3\,4\,7 \\ 3\,7\,6\,7\,6\,7\,2 & & 3\,4\,6\,4\,7\,6\,2 \\ 4\,5\,2\,2\,5\,6\,5 & & 2\,7\,5\,5\,6\,2\,5. \end{array}$$

Now the powers of U are

$$U = (347562), \quad U^2 = (376)(452), \quad U^3 = (35)(46)(72),$$
$$U^4 = (367)(425), \quad U^5 = (326574), \quad U^6 = 1.$$

Thus the entire group leaving the polyhedron, and in particular the polyhedron of 7 hexagons, on the torus invariant is

$$(S_1, S_2, \cdots, S_7)(1, T)(1, U, U^2) \equiv G_{42},$$

as is well known.*

4. **Triple systems of order 9 with double couples.** The triple system obtained from

$$\left\| \begin{array}{ccc} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{array} \right\| \equiv \begin{array}{c} 1\,1\,1\,1\,2\,2\,2\,3\,3\,3\,4\,7 \\ 2\,4\,5\,6\,4\,6\,5\,5\,4\,6\,5\,8 \\ 3\,7\,9\,8\,9\,7\,8\,7\,8\,9\,6\,9 \end{array}$$

has a group which is doubly transitive, since 1, 2 may be replaced by any two other letters, say 5 and 8:

$$\left\| \begin{array}{ccc} 5 & 8 & 2 \\ 1 & 4 & 7 \\ 9 & 3 & 6 \end{array} \right\|.$$

Or we have the substitution (154)(283)(796) which leaves the triple

* Brahana, *Regular maps on an anchor ring*, American Journal of Mathematics, vol. 48 (1926), pp. 225–240.

system invariant. Moreover the substitutions which leave 1, 2 invariant also leave 3 invariant and form a group of order 6:

1, (47)(58)(69), (456)(798), (465)(789), (59)(67)(48), (68)(49)(57).

Hence the group which leaves the triple system invariant is a G_{432} of order $9\cdot 8\cdot 6=432$, since a group on n elements which is k-fold transitive has the order equal to the product of $n(n-1)\cdots(n-k+1)$ and the order of the group which leaves k of the elements fixed. The G_{432} is also the holomorph of the non-cyclic abelian group Γ_9 of order 9:

$$\Gamma_9(S_1,S_2,S_3,S_4),$$

in which

$$S_1 = 123\cdot 456\cdot 789,$$
$$S_2 = 147\cdot 258\cdot 369,$$
$$S_3 = 159\cdot 267\cdot 348,$$
$$S_4 = 168\cdot 249\cdot 357.$$

Two triple systems Δ_9^3 without a common triple form a closed surface with three double vertices:

$$\left\|\begin{matrix}1&2&3\\4&5&6\\7&8&9\end{matrix}\right\| \equiv \begin{matrix}1&1&1&1&2&2&2&3&3&3&4&7\\2&4&5&6&4&5&6&4&5&6&5&8\\3&7&9&8&9&8&7&8&7&9&6&9\end{matrix}$$

$$\left\|\begin{matrix}1&2&4\\3&5&9\\7&6&8\end{matrix}\right\| \equiv \begin{matrix}1&1&1&1&2&2&2&3&3&4&4&7\\2&3&5&6&3&5&7&4&5&5&8&6\\4&7&8&9&8&6&9&6&9&7&9&8.\end{matrix}$$

The surface can be disected and mapped on a rectangle, Fig. 2, then, as in Fig. 1b, on a cylinder, and finally on a torus.

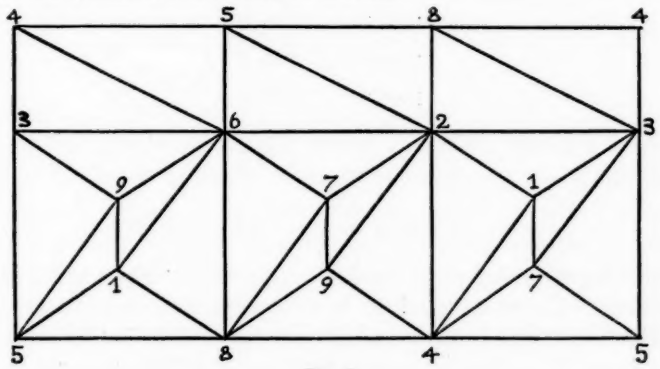


FIG. 2

There is to be noticed this peculiarity in the polyhedral configuration for $n=9$, that the three vertices 1, 7, 9 are double vertices, each with two distinct cycles of triangles. By Euler's formula

$$k = V - S + F = (6 + 3\cdot 2) - 36 + 24 = 0,$$

so that $2-k=2$ is the order of connectivity.

From the rectangular figure it is apparent that the cylindrical configuration, with the edges $\overline{534}$ joined, is left invariant by the cyclic substitutions (rotations around axis of cylinder)

$$(458)(236)(197) = T,$$

moreover by

$$S_1 = (28)(35)(46)(79),$$
$$S_2 = (17)(25)(34)(68),$$
$$S_3 = (19)(38)(56)(24).$$

Now

$$S_1S_3 = T = (236)(458)(719),$$
$$S_1S_2 = (263)(485)(917) = T^2,$$
$$S_2S_3 = (791)(263)(485) = T^2.$$

Hence there is a non-cyclic group G_6 of order 6 $(1, T, T^2, S_1, S_2, S_3)$ which leaves the closed surface invariant.

5. **Triple systems of even order with double couples.** The simplest case of this sort is for $n=6\mu$, $\mu=1$, or $n=6$, with $x=2\mu(6\mu-1)=10$.

Such a system is for example

1 2 3	2 3 5
1 3 4	3 4 6
1 4 5	4 5 2
1 5 6	5 6 3
1 6 2	6 2 4

which form a closed surface as shown when mapped on a plane, Fig. 3. It is unilateral and can be mapped on a projective plane. The formula gives

$$V - S + F = 10 + 6 - 15 = 1 = k$$

or $2-k=1$ for the order of connectivity. The substitutions which leave the surface in its composition of triples invariant are the cycles of order 3 which rotate Fig. 3 about its center. Such a cycle is $123\cdot 465$, $132\cdot 456$. But every triple may be chosen at the center, thus we have $10\cdot 2$ such substitutions.

Then there are cycles of order 5 which leave one point invariant, for example 1, i.e., (23456), (24635), (25364), (26543), or $6 \cdot 4 = 24$ such substitutions. Finally there are 15 substitutions of order 2, each forming a spear head like (13) (24), or (16) (34), (16) (25). Hence there are altogether $1+15+20+24 = 60$ substitutions which leave the triple system invariant. *If we join the centers of the five triangles around each vertex we obtain the unilateral surface mapped into* 6 *pentagons, i.e., a unilateral polyhedron of* 6 *pentagons on the projective plane and* 10 *vertices is left invariant by the icosahedral group.*

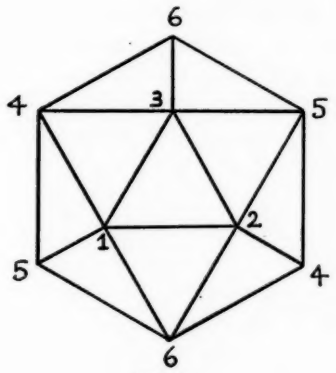


FIG. 3

Considering the remaining 10 triples, they form also a triple system with double couples

1 2 4	2 4 3
1 4 6	4 6 5
1 6 3	6 3 2,
1 3 5	3 5 4
1 5 2	5 2 6

which is left invariant by the same icosahedral group. As can easily be verified, each triple is common to six distinct triple systems. With every triple system is associated the complementary triple system formed by the other 10 of the total of 20 triples. In a triple system and its complement all 20 triples are contained. *Hence there are altogether* 12 *distinct triple systems and in this complex each triple is contained just* 6 *times.*

III. Multiple systems

1. **Systems Δ_n^k.** *n elements 1, 2, 3, $\cdots$, n form $\binom{n}{k}$ k-tuples and $\binom{n}{k-1}$ $(k-1)$-tuples. It is proposed to form a system Δ_n^k of k-tuples with the property that it contains each $(k-1)$-tuple just once.*

Let x be this number; then since every k-tuple contains k $(k-1)$-tuples

$$kx = \binom{n}{k-1}, \text{ or } x = \frac{n(n-1)(n-2)\cdots(n-k+2)}{1\cdot 2\cdot 3\cdots(k-1)k},$$

which must be an integer. The condition that this be an integer is however not sufficient. Let y be the number of k-tuples of Δ_n^k containing the letter 1. Then all $(k-1)$-tuples of Δ_n^k containing the letter 1 are in these k-tuples. A $(k-2)$-tuple formed from the remaining $n-1$ letters cannot appear in more than one of the y k-tuples. On the other hand each of these $(k-2)$-tuples is contained in the y k-tuples. Thus in each of the y k-tuples, there are $\binom{k-1}{k-2}$ $(k-2)$-tuples. Hence

$$y\binom{k-1}{k-2} = \binom{n-1}{k-2},$$

or

$$y = \frac{(n-1)(n-2)\cdots(n-k+2)}{(k-1)(k-2)\cdots 2\cdot 1}$$

is the number of k-tuples of Δ_n^k with a given letter just once in each of the y k-tuples. For a Δ_n^k system y must be an integer, which is another necessary condition. Next we might find the number of k-tuples that each contain a given couple ik and so forth.

2. Δ_n^4. To illustrate this, take $k=4$; then

$$x = \frac{n(n-1)(n-2)}{1\cdot 2\cdot 3\cdot 4}, \qquad y = \frac{(n-1)(n-2)}{3\cdot 2}.$$

In this case $n-2$ elements which form quadruples with ik must form $\frac{1}{2}(n-2)$ couples which with ik form all possible distinct triples each containing ik. From this follows that n must be even. Thus $n=2m$, and $y=\frac{1}{3}(m-1)\cdot(2m-1)$, and either $m=3\mu+1$, or $m=\frac{1}{2}(3\mu+1)$. In the latter case $\mu=2\gamma+1$ and $m=3\gamma+2$.

For quadruple systems of order n we have therefore either $n=6\mu+2$, or $n=6\mu+4$, and the number of quadruples in Δ_n^4 is accordingly

$$x = \tfrac{1}{2}(3\mu+1)(6\mu+1)\mu, \quad \text{or} \quad x = \tfrac{1}{2}(3\mu+2)(2\mu+1)(3\mu+1).$$

In such a system each element occurs $\mu(6\mu+1)$ or $(2\mu+1)(3\mu+1)$ times.

3. **Quadruple systems of lowest order Δ_8^4.** Next to the quadruple system on four elements which consists of one quadruple 1234, containing four triples 123, 124, 134, 234, and which is left invariant by the symmetric group on four elements, the quadruple systems of lowest order can be formed with 8 elements, such that each system contains 14 quadruples and each possible triple once, and every couple three times, and each letter 7 times.

It is not difficult to devise a graphical scheme to establish such quadruple systems. One is to number the vertices of a cube from 1 to 8 in any order. Then the quadruples of the system consist first of 6 faces, second of the 6 rectangles each determined by two opposite edges, third by the two tetrahedrons inscribed in the cube:

1 4 1 2 1 3 1 3 2 1 1 2 2 1
2 6 5 3 2 4 2 5 3 4 3 5 4 3
3 7 6 4 6 5 4 6 6 5 4 7 5 7
5 8 7 8 8 7 7 8 7 8 6 8 6 8 .

The problem now is to determine the order of the group of degree 8 which will permute the quadruples of the system, in which a quadruple is defined as a set of four numbers irrespective of the order in which they are written. In the first place it is evident that the octahedral group written on 8 letters leaves the system invariant and hence is a subgroup of the group in question. Moreover, as any three letters on the cube can be replaced by any three other letters and the remaining one so distributed over the remaining vertices as to give rise to the same quadruple systems, the group will be triply transitive. The substitutions of the group which leave 3 letters unchanged, for instance 123, also must leave 5 unchanged. These are 1, (46)(78), (47) (68), (48) (67), which form a subgroup of order 4. Hence according to a well known theorem that the order of a group of degree n which is k-fold transitive is equal to the product of $n(n-1) \cdots (n-k+1)$ and the order of the group which leaves k elements unchanged, the order of the group of the quadruple system is

$$8 \cdot 7 \cdot 6 \cdot 4 = 1344.$$

This order may also be obtained by establishing the number of distinct quadruple systems. As every triple must be in a system,

1234, 1235, 1236, 1237, 1238

must belong to different quadruple systems. The problem then is to find how many quadruple systems can be formed with each of those quadruples fixed. Considering 1235, for example, we can form 24 permutations with

4678, each of which with 1235 defines a quadruple system. But any two successive transpositions of different letters, like (47)(68), leaves the quadruple system invariant. As there are 4 such double transpositions, there are 6 distinct substitutions, (48), (67), (47), (68), (78), (46), which will give three distinct quadruple systems. Thus either (46), (47), (48), or (78), (68), (67), give three distinct systems; likewise the product of any two, i.e., (46)(47), (47)(48), (48)(46), three more, i.e. altogether 6 quadruple systems associated with 1235. Thus 4876 is obtained by the substitution (78) (67) or (47) (68)·(46) (47) from 4768; likewise 4687 from 4768 by (67) (87) or (678). As there are altogether 8! substitutions and 30 distinct quadruple systems, each quadruple must be invariant in 8!:30 = 1344 substitutions, which form the above G_{1344}.

As to the structure of this group it contains a G_8:

(1)
(12)(35)(47)(68)
(13)(25)(46)(78)
(15)(23)(48)(67)
(18)(26)(37)(45)
(16)(28)(34)(57)
(17)(24)(38)(56)
(14)(27)(36)(58) ;

a G_8':

(1253)(48)
(15)(23)
(1352)(48)
(12)(35)
(23)(48)
(13)(25)
(15)(48) ;

a G_7:

(1234587) cyc. ;

a G_3:

(235)(478) cyc. ;

moreover the subgroups

$$G_{64} = G_8 \cdot G_8', \quad G_8 \cdot G_7 = G_{56}, \quad G_8 \cdot G_3 = G_{24},$$

$$G_7 \cdot G_3 = G_{21}, \quad G_8 \cdot G_7 \cdot G_3 = G_{168} \quad \text{(not simple)}.$$

4. Configuration of two quadruple systems without a common quadruple. If we interpret the eight elements as eight points in space of four dimensions, every quadruple will be represented by a tetrahedron. Two quadruple systems without a common quadruple can be joined so that each face (triple) of the fourteen tetrahedrons in one system will coincide with the same triple-face of the tetrahedron in the second system in which it occurs. In this manner we obtain a closed hypersurface P_{28} in S_4 bounded by 28 tetrahedral cells.

If in an S_n there are, for a polyhedron, r_0 vertices, r_1 edges, r_2 faces, r_3 cells, r_4 four-spaces bounding, etc., then for an Eulerian polyhedron

$$1 - r_0 + r_1 - r_2 + r_3 + \cdots + (-1)^{n-1} r_{n-1} = 0 .$$

Here r_{n-1}, the number of $(n-1)$-spaces bounding the polyhedron, is 1. Now for our 28-cell in S_4 we have

$$1 - 8 + 28 - 56 + 28 - 1 = -8.$$

Hence the 28-cell in S_4 does not belong to the ordinary polytope.

The cells of a P_{28} may be represented for example by

$$Q_{14}\begin{cases} 1\,1\,1\,1\,1\,1\,1\,2\,2\,2\,2\,3\,3\,4 \\ 2\,2\,2\,3\,3\,4\,5\,3\,3\,4\,5\,4\,5\,6 \\ 3\,4\,6\,4\,7\,5\,6\,4\,6\,5\,7\,5\,6\,7 \\ 5\,7\,8\,6\,8\,8\,7\,8\,7\,6\,8\,7\,8\,8 \end{cases},$$

$$Q_{14}'\begin{cases} 1\,1\,1\,1\,1\,1\,1\,3\,4\,2\,2\,2\,2\,3 \\ 2\,3\,5\,3\,4\,2\,2\,4\,5\,3\,5\,3\,4\,6 \\ 4\,5\,6\,4\,6\,3\,7\,5\,7\,5\,6\,4\,6\,7 \\ 5\,7\,8\,8\,7\,6\,8\,6\,8\,8\,7\,7\,8\,8 \end{cases},$$

so that

$$P_{28} = Q_{14} + Q_{14}',$$

in which Q_{14} and Q_{14}' are two quadruple systems without a common quadruple. The problem is to find the group of degree 8 which will leave P_{28} invariant. There are substitutions which leave any two elements in Q_{14} and the same two elements in Q_{14}' fixed and permute Q_{14} and Q_{14}'; for example, $S=(25)\,(34)\,(18)$ which leaves (1) and (7), $T=(37)\,(46)\,(58)$ which leaves (1) and (2), $U=(13)\,(68)\,(27)$ which leaves (4) and (5) invariant and also permutes Q_{14} and Q_{14}'. The product of any such two substitutions is a substitution which leaves each Q_{14} and Q_{14}' invariant, for example

$$ST = (2847365), \quad SU = (134)(257).$$

Thus for each element there is a cyle of order 7 which leaves this element and each Q_{14} and Q_{14}' invariant. It is seen that the substitutions of the group which leave P_{28} invariant, are of two kinds: those that permute Q_{14} and Q_{14}' and those that leave both Q_{14} and Q_{14}' invariant. Another substitution which leaves (1) and (2) fixed and permutes Q_{14} and Q_{14}' is $W=(368745)$, for which $W^2=(384)\,(675)$, $W^3=(37)\,(46)\,(58)=T$. $SUT=(17285364)$ is a substitution of order 8 which permutes Q_{14} and Q_{14}'. $(SUT)^2=(1256)\,(3478)$ and $(SUT)^4=(15)\,(26)\,(37)\,(48)$ are substitutions which transform Q_{14} and Q_{14}' into themselves. Hence there are operators of order 2, 3, 4, 7 which transform Q_{14} and Q_{14}' into themselves, and operators of order 2, 6, 8 which permute Q_{14} and Q_{14}'.

The group is again triply transitive and the substitutions which leave three elements, say (1), (2), (3), invariant cannot also leave (5) and (6) invariant (outside of identity) hence 5 and 6 must be permuted, likewise 7 and 8, but these assumptions lead to contradictions, so that the only substitution which leaves 1, 2, 3, in $Q_{14}+Q_{14}'$, invariant is identity. Hence the group leaving P_{28} invariant is of order $8\cdot 7\cdot 6\cdot 1=336$. The odd substitutions of this group permute Q_{14} and Q_{14}' while the even substitutions form the simple G_{168} which leaves both Q_{14} and Q_{14}' invariant. There is

$$G_{336}=(1,C,C^2,\cdots)(1,A,A^2,\cdots)(1,B,B^2,\cdots),$$

where

$$A=(285743),$$
$$B=(1234567),$$
$$C=(17285364).$$

The orders of A and C are 6 and 8 respectively. The product $C^\alpha A^\beta$ leaves Q_{14} and Q_{14}' invariant, when $\alpha+\beta=2m$. As $A^6=1$, $C^8=1$, we may solve this equation for $\alpha+\beta\leqq 12$, $\alpha\leqq 8$, $\beta\leqq 6$, and find easily 24 solutions, including identity.

Hence there are 24 *substitutions of the form* $C^\alpha A^\beta$ *which leave* Δ_4^8 *and* $\Delta_4'^8$ *invariant, consequently leave the configuration* $\Delta_4^8\cdot\Delta_4^8$, *or the polyhedron* P_{28} *in* S_4 *invariant.*

That configurations admitting finite groups are not limited to the types Δ_n^k will appear from the following example.

5. **A certain quadruple system on 16 elements.** It is proposed to construct a quadruple system on 16 elements 1, 2, $\cdots$, 16 such that every element occurs, just 6 times, i.e., in six quadruples of the system.

Let x be the number of quadruples; then there are $4x$ elements in the system. But as every element occurs in six quadruples, it follows that

$$\frac{4x}{6} = 16,$$

and $x = 24$. Thus such a system must have 24 quadruples. Such a system is for example

1· 2· 3· 4	1· 5· 4· 8	5· 9·12· 8	1· 4·16·13
5· 6· 7· 8	2· 6· 7· 3	9·13·16·12	1· 2·14·13
9·10·11·12	8·12·11· 7	10·14·15·11	1· 5· 9·13
13·14·15·16	12·16·15·11	6·10·11· 7	2· 3·15·14
1· 5· 6· 2	5· 9·10· 6	7·11·15· 3	6·10·14· 2
8· 7· 3· 4	9·13·14·10	3·16·15· 4	8· 4·16·12.

There are exactly 32 couples each of which occurs exactly 3 times in the system. The 24 quadruples may be grouped into 8 octuples:

1· 2· 3· 4	5· 6· 7· 8	9·10·11·12	13·14·15·16
5· 6· 7· 8	9·10·11·12	13·14·15·16	1· 2· 3· 4
8· 4· 1· 5	5· 6· 2· 1	6·10·14· 2	8· 7· 3· 4
12·16·13· 9	10·14·13· 9	7·11·15· 3	12·11·15·16,

so that every octuple contains exactly 6 of the quadruples of the system. But the 48 quadruples thus obtained are two by two equal, so that there are only the 24 original distinct quadruples. *Now between the number of elements, couples, quadruples and octuples exists the identity*

$$1 - 16 + 32 - 24 + 8 - 1 = 0$$

which is Euler's formula for the 8-cell, or cube in four dimensions. Geometrically this follows from the following consideration. The 16 elements may be represented by 16 generic points in euclidean four-space. By a continuous deformation of space, more precisely by a transformation of at least the second order

$$\rho x_i' = \frac{\sum_{k,l=1}^{4} a_{ikl}x_k x_l + d_{ik}x_k + b_i}{\sum_{k,l=1}^{4} a_{kkl}x_k x_l + C_{ik}x_k + d} \qquad (i = 1,2,3,4),$$

depending on 74 effective constants, 16 generic points, determined by $4 \cdot 16 = 64$ constants, may always be transformed into the 16 vertices of an 8-cell. Thus when they are transformed in the order as indicated by the above quadruple system, the elements, couples, quadruples, and octuples of the system form the vertices, edges, faces and cubes of an 8-cell in four dimensions. The octahedral group attached to each cube of the 8-cell leaves this cube invariant and induces a definite transformation of the remaining cubes among themselves, thus transforming the whole 8-cell into itself. With each cube there is associated such an octahedral group. Hence there are $8 \cdot 24 = 192$ movements of the 8-cell into itself. Or the group associated with the quadruple system is a G_{192}, which is well known as the polyhedral group which leaves the 8-cell invariant.

Every permutation of the 16 elements marked off in quadruples as in

$$1\ 2\ 3\ 4 \cdot 5\ 6\ 7\ 8 \cdot 9\ 10\ 11\ 12 \cdot 13\ 14\ 15\ 16$$

and formed into a quadruple system as induced by this grouping gives rise to a quadruple system which is invariant in a G_{192}.

6. **Quintuple and higher systems.** By a similar procedure, as in Δ_n^4 it is found that the lowest possible non-trivial quintuple system is a Δ_{11}^5 with 66 quintuples. In such a system every letter occurs 30 times, every couple 12 times, every triple 4 times, every quadruple once.

Professor Carmichael shows that the group which leaves this system invariant is a four-fold transitive Mathieu group of degree 11 and order $11 \cdot 10 \cdot 9 \cdot 8$.

The existence proofs for quintuple systems for $n > 11$, and in general of k-tuple systems, seem to be difficult, likewise the solution of the problem of complete sets of k-tuple systems for possible values of n. In any case these problems are closely connected with group theory.

As in the case of the 8-cell in S_4, other conditions than the existence of each $(k-1)$-tuple in a set of k-tuples may be imposed on n elements. We get accordingly different configurations and groups in which they are invariant.

UNIVERSITY OF ILLINOIS,
URBANA, ILL.

ON THE EXPANSION OF ANALYTIC FUNCTIONS OF THE COMPLEX VARIABLE IN GENERALIZED TAYLOR'S SERIES*

BY

D. V. WIDDER†

1. Introduction. In a previous paper‡ in these Transactions the representation of real functions by generalized Taylor's series was discussed. The methods there employed are not applicable to the corresponding representation of functions of the complex variable. In the present paper new methods are developed, and the representation of an arbitrary function of the complex variable is obtained.

Let us denote the function to be represented by $f(z)$ and the functions in terms of which the representation is to be made by

$$u_0(z),\ u_1(z),\ \cdots. \tag{1}$$

As in the previous paper we begin by a sort of "normalization"§ of this sequence, replacing it by a sequence

$$g_0(z),\ g_1(z),\ \cdots, \tag{2}$$

where the function $g_n(z)$ is a linear combination of $u_0(z), u_1(z), \cdots, u_n(z)$, and has a zero of order n exactly at the origin. The possibility of such a normalization imposes, of course, a restriction on the sequence (1) which we shall specify in §2. Indeed, a number of other conditions must be imposed, as one would expect, in order to insure the representation of an arbitrary analytic function $f(z)$ by means of linear combinations of the functions (1). These conditions are most conveniently expressed in terms of the set (2), and amount briefly to demanding that they be analytic and that $|1-n!g_n(z)/z^n|$ should decrease at least as rapidly as $1/n$ when n becomes infinite. It is then found that the region of convergence of a series

$$\sum_{n=0}^{\infty} c_n g_n(z) \tag{3}$$

---

* Presented to the Society, September 6, 1927; received by the editors in March, 1928.

† National Research Fellow in Mathematics.

‡ D. V. Widder, *A generalization of Taylor's series*, these Transactions, vol. 30 (1928), p. 126.

§ This process corresponds to orthogonalizing and normalizing the sequence of prescribed functions by the Schmidt method, when the method of approximation employed is that of least squares.

is circular, and that an analytic function $f(z)$ can be represented in a series of this type inside a circle with center at the origin and reaching out to the nearest singularity of $f(z)$, just as in the case of Taylor's series. These facts we shall prove in §§2 and 3.

In §4 we compare the result just stated with one of G. D. Birkhoff* on generalized Taylor's series. We show that the latter result demands implicitly that the functions $v_n(z)$ of Birkhoff* should have exactly n zeros inside the circle $|z|<1$. In that respect the function is like the function $g_n(z)$ of the present paper, which has n zeros coincident at the origin and no others in the unit circle if n is sufficiently large. But it is found that the conditions imposed on the mode of decrease of $|h_n(z)|$ in the present paper are of such a nature that a class of expansions arise which can not be treated by Birkhoff's theorem. When the expansion can be treated by both methods a comparison of the two methods of determining the coefficients (for a common function $f(z)$) yields an interesting identity.

Finally, we cite several other papers concerned with problems related to our own, but which do not treat the case considered by us:

S. Kakeya, *On the expansion of analytic function*, Proceedings of the Physico-Mathematical Society of Japan, (3), vol. 2 (1920), p. 96.

Y. Okada, *On a certain expansion of analytic function*, Tôhoku Mathematical Journal, vol. 22 (1923), p. 325.

J. L. Walsh, *On the expansion of analytic functions in series of polynomials*, these Transactions, vol. 26 (1924), p. 155.

S. Izumi, *On the expansion of analytic function*, Tôhoku Mathematical Journal, vol. 28 (1927), p. 97.

R. F. Graesser, *A certain general type of Neumann expansion and expansions in hypergeometric functions*, American Journal of Mathematics, vol. 49 (1927), p. 577.

J. L. Walsh, *On the expansion of analytic functions in series of polynomials and in series of other analytic functions*, these Transactions, vol. 30 (1928), p. 307.

We wish to emphasize the fact that the series of the present paper converge throughout the circular region of analyticity of the function $f(z)$ and represent $f(z)$ there, a property not shared by the series of Izumi and Graesser, for example. This property is retained from Taylor's series in the generalization, and is also shared by the series of Birkhoff and Walsh.

2. **The region of convergence.** Let the functions (1) and $f(z)$ be analytic

* G. D. Birkhoff, *Sur une généralisation de la série de Taylor*, Comptes Rendus, vol. 164, p. 942. For a statement of the theorem see §4 of the present paper.

in the circle $|z| \leqq R$. Denote by $s_n(z)$ a linear combination of the first $n+1$ functions (1):

$$s_n(z) = c_0u_0(z) + c_1u_1(z) + c_2u_2(z) + \cdots + c_nu_n(z),$$

and determine the constants of combination $c_0, c_1, \cdots, c_n$ in such a way that $f(z)-s_n(z)$ shall have a zero of order $n+1$ at $z=0$. This is possible if the Wronskian

$$W_n(z) = W[u_0(z), u_1(z), \cdots, u_n(z)] = \begin{vmatrix} u_0(z) & u_1(z) & \cdots & u_n(z) \\ u_0'(z) & u_1'(z) & \cdots & u_n'(z) \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ u_0^{(n)}(z) & u_1^{(n)}(z) & \cdots & u_n^{(n)}(z) \end{vmatrix}$$

is different from zero at $z=0$. We assume that $W_n(z) \neq 0$ for $|z| \leqq R$ and for all* n. The computations of the previous paper are evidently applicable here, so that $s_n(z)$ takes the form

$$s_n(z) = L_0f(0)g_0(z) + L_1f(0)g_1(z) + \cdots + L_nf(0)g_n(z),$$

where

$$g_n(z) = \frac{1}{W_n(0)} \begin{vmatrix} u_0(0) & u_1(0) & \cdots & u_n(0) \\ u_0'(0) & u_1'(0) & \cdots & u_n'(0) \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ u_0^{(n-1)}(0) & u_1^{(n-1)}(0) & \cdots & u_n^{(n-1)}(0) \\ u_0(z) & u_1(z) & \cdots & u_n(z) \end{vmatrix},$$

$$L_nf(z) = \frac{W[u_0(z), u_1(z), \cdots, u_{n-1}(z), f(z)]}{W_{n-1}(z)} \quad (n = 0,1,2, \cdots).$$

Since the functions $g_n(z)$ are linear combinations of the functions (1), it is clear that $L_nf(z)$ may equally well be expressed in terms of $g_n(z)$:

$$L_nf(z) = \frac{W[g_0(z), g_1(z), \cdots, g_{n-1}(z), f(z)]}{W[g_0(z), g_1(z), \cdots, g_{n-1}(z)]}.$$

---

* J. L. Walsh has called the author's attention to the fact that if we admit the rearrangement of the order of the functions in the sequence (1), then it is only necessary to assume here that $W_n(0) \neq 0$ for *n sufficiently large*. The proof of this fact becomes evident by use of a lemma of Walsh to be found in the American Journal of Mathematics, vol. 42 (1920), p. 93.

It is seen that the functions $g_n(z)$ have the properties enunciated in the introduction:

$$g_n^{(k)}(0) = 0 \qquad (k = 0,1,2, \cdots, n-1), \qquad g_n^{(n)}(0) = 1.$$

Since $g_n(z)$ is analytic for $|z| \leqq R$, it must have the form

$$g_n(z) = z^n[1 + h_n(z)]/n!,$$

where $h_n(z)$ is analytic in the same circle and vanishes at the center. We further restrict the sequence (2) by the

CONDITIONS A: (a) *The functions $g_n(z)$ are analytic for $|z| \leqq R$;* (b) *$g_n(z) = z^n[1+h_n(z)]/n!$, where $h_n(z)$ is analytic for $|z| \leqq R$ and vanishes at $z=0$;* (c) *a constant M independent of n exists such that $|h_n(z)| \leqq M/(n+1)$ for $|z| \leqq R$.*

We can now state

THEOREM I. *Let the functions $g_n(z)$ satisfy Conditions* A *in $|z| \leqq R$. Then if the series*

$$\sum_{n=0}^{\infty} c_n g_n(z) \tag{3}$$

converges for one value $z=\xi$, $|\xi| \leqq R$, it converges uniformly for $|z| \leqq r$ for any $r < |\xi|$ to an analytic function.

To prove this, note first that the convergence of (3) for $z=\xi$ implies the existence of a constant N independent of n such that

$$|c_n g_n(\xi)| \leqq N,$$

so that a dominant series for (3) can be obtained at once:

$$\sum_{n=0}^{\infty} c_n g_n(z) \ll N \sum_{n=0}^{\infty} \frac{|g_n(z)|}{|g_n(\xi)|}.$$

By Conditions A,

$$|g_n(z)| \leqq \frac{|z|^n}{n!}\left(1 + \frac{M}{n+1}\right), \quad |z| \leqq R,$$

and

$$|g_n(\xi)| = \frac{|\xi|^n}{n!}\left|1 + h_n(\xi)\right| \geqq \frac{|\xi|^n}{n!}\left(1 - \frac{M}{n+1}\right).$$

Hence

$$\sum_{n=m}^{\infty} c_n g_n(z) \ll N \sum_{n=m}^{\infty} \frac{|z|^n}{|\xi|^n} \frac{\left(1 + \frac{M}{n+1}\right)}{\left(1 - \frac{M}{n+1}\right)} \ll N \sum_{n=m}^{\infty} \frac{r^n}{|\xi|^n} \frac{\left(1 + \frac{M}{n+1}\right)}{\left(1 - \frac{M}{n+1}\right)}.$$

Here m is chosen so large that $m+1>M$. If $r<|\xi|$, (3) clearly converges uniformly for $|z|\leqq r$, and the theorem is established.

One may show, just as for power series, that if (3) diverges for a value $z=\xi$ with $|\xi|<R$, it diverges for $|\xi|\leqq|z|\leqq R$. It follows then that (3) converges nowhere except at $z=0$, or converges for $|z|<r<R$ and diverges for $r<|z|\leqq R$, or converges for $|z|\leqq R$. It is easy to construct examples to show that all three possibilities actually arise. It should be pointed out that in the third case we have not proved that the true region of convergence is circular, for (3) may converge outside of the circle $|z|\leqq R$ in which Conditions A hold. In case the region of convergence is a circle we define the circle as the *circle of convergence* and its radius as the *radius of convergence.*

THEOREM II. *The radius of convergence ρ of* (3), *if it is less than R, is given by the equation*

$$\frac{1}{\rho}=\overline{\lim_{n=\infty}}(|c_n|/n!)^{1/n}. \tag{4}$$

For, if (4) holds then to every positive ϵ there corresponds an integer m such that

$$|c_n|<(1+\epsilon)^n n!/\rho^n, \qquad n\geqq m,$$

and by virtue of Conditions A we have

$$\sum_{n=m}^{\infty}c_n g_n(z)\ll\sum_{n=m}^{\infty}\frac{(1+\epsilon)^n}{\rho^n}|z|^n\left(1+\frac{M}{n+1}\right).$$

Hence (3) converges for $|z|<\rho$. It diverges for $|z|>\rho$, for by (4) it follows that

$$|c_n|>\frac{(1-\epsilon)^n}{\rho^n}n!, \qquad \epsilon>0,$$

for an infinite number of integers n. For these integers

$$|c_n g_n(z)|>\frac{|z|^n}{\rho^n}\left(1-\frac{M}{n+1}\right)(1-\epsilon)^n,$$

and if $|z|>\rho$ and $(1-\epsilon)/|z|>\rho$ the general term of (3) can not approach zero as n becomes infinite. Hence ρ must be the radius of convergence. A similar proof shows that if $\overline{\lim}\,(|c_n|/n!)^{1/n}=\infty$ then $\rho=0$, and we have the first case mentioned above.

3. **The representation of an arbitrary analytic function.** Let us suppose that $f(z)$ is analytic in the circle $|z|<\rho\leqq R$. We shall show that the series

$$\sum_{n=0}^{\infty} L_n f(0) g_n(z) \tag{5}$$

converges and represents $f(z)$ in that circle provided that the functions $g_n(z)$ satisfy Conditions A in $|z| \leqq R$. To prove this we first prove that (5) converges for $|z| < \rho$. This is done by use of the following formula:

$$f^{(n)}(0) = L_0 f(0) g_0^{(n)}(0) + L_1 f(0) g_1^{(n)}(0) + \cdots + L_{n-1} f(0) g_{n-1}^{(n)}(0) + L_n f(0), \tag{6}$$

taken from the previous paper.* A moment's consideration will show that the proof there given for the real variable applies equally well here. By its use we can determine the mode of increase of $|L_n f(0)|$. We first transform the formula so as to involve the functions $h_n(z)$ instead of $g_n(z)$. By the definition of $h_n(z)$ it is clear that

$$g_n^{(k)}(z) = \sum_{p=0}^{k} \binom{k}{p} \left(\frac{z^n}{n!}\right)^{(p)} (1 + h_n(z))^{(k-p)},$$

where

$$\binom{k}{p} = \frac{k!}{p!(k-p)!}.$$

For formula (6) we shall want $k > n$. Then

$$g_n^{(k)}(0) = \binom{k}{n} (1 + h_n(z))^{(k-n)} \Big|_{z=0} = \binom{k}{n} h_n^{(k-n)}(0).$$

Hence (6) reduces to

$$L_n f(0) = f^{(n)}(0) - \sum_{p=0}^{n-1} \binom{n}{p} h_p^{(n-p)}(0) L_p f(0). \tag{7}$$

Let r be an arbitrary positive constant less than ρ. The analyticity of $f(z)$ in $|z| < \rho$ implies the existence of a constant L independent of n such that

$$|f^{(n)}(0)| \leqq \frac{Ln!}{r^n}, \qquad |z| \leqq r. \tag{8}$$

Since $h_n(z)$ is analytic in $|z| \leqq r$ and satisfies the inequality

$$|h_n(z)| \leqq M/(n+1), \quad |z| \leqq R,$$

it follows that

$$|h_n^{(k)}(0)| \leqq \frac{M}{n+1} \frac{k!}{r^k}. \tag{9}$$

Now

* Loc. cit., p. 131.

$$L_0 f(0) = f(0), \qquad |L_0 f(0)| \leqq L,$$
$$L_1 f(0) = f'(0) - L_0 f(0) h_0'(0),$$
$$|L_1 f(0)| \leqq L/r + LM/r = L(1+M)/r.$$

Proceeding in this way one is led to induce the inequality

$$|L_n f(0)| \leqq L(M+1)(M+2)\cdots(M+n)/r^n.$$

To establish this result we may assume it true for $n=0, 1, 2, \cdots, m-1$, and show as a consequence that it is also true for $n=m$. By (7), (8), (9), and (10)

$$|L_m f(0)| \leqq |f^{(m)}(0)| + \sum_{p=0}^{m-1} \binom{m}{p} |L_p f(0)| \cdot |h_p^{(m-p)}(0)|$$
$$\leqq \frac{Lm!}{r^m} + \sum_{p=0}^{m-1} \binom{m}{p} \frac{L}{r^p} (M+1)(M+2)\cdots(M+p) \frac{M}{p+1} \frac{(m-p)!}{r^{m-p}}.$$

It is easily verified that the expression on the right of this inequality is precisely

$$L(M+1)(M+2)\cdots(M+m)/r^m, \tag{10}$$

so that the induction is complete.

We can now establish the convergence of (5) for $|z|<\rho$. For, we have

$$\sum_{n=0}^{\infty} L_n f(0) g_n(z) \ll L \sum_{n=0}^{\infty} \frac{(M+1)(M+2)\cdots(M+n)}{r^n} \frac{|z|^n}{n!} \left(1 + \frac{M}{n+1}\right).$$

Simple tests show that the dominant series converges for $|z|<r$. Since r was arbitrary it follows that (5) converges absolutely for $|z|<\rho$. It is equally evident that (5) converges uniformly for $|z| \leqq r$ if $r<\rho$. Consequently the sum of the series, which we denote by $\phi(z)$, must be an analytic function in the circle $|z|<\rho$. Moreover, the series may be differentiated term by term as often as desired by a familiar theorem of Weierstrass.

It will now be established that $\phi(z)=f(z)$ for $|z|<\rho$. Setting $z=0$ it is seen immediately that $\phi(0)=f(0)$. Differentiating (5) term by term k times, it is clear that

$$\phi^{(k)}(0) = \sum_{n=0}^{\infty} L_n f(0) g_n^{(k)}(0) = \sum_{n=0}^{k} L_n f(0) g_n^{(k)}(0).$$

But by formula (6)

$$f^{(k)}(0) = \sum_{n=0}^{k} L_n f(0) g_n^{(k)}(0).$$

Hence

$$f^{(k)}(0) = \phi^{(k)}(0) \qquad (k = 0,1,2, \cdots).$$

Since $f(z)$ and $\phi(z)$ are both analytic in the neighborhood of the origin, thev must coincide throughout their region of analyticity, so that

$$f(z) = \sum_{n=0}^{\infty} L_n f(0) g_n(z), \qquad |z| < \rho.$$

We have thus proved not only that the expansion is possible but also that it is unique (the functions $g_n(z)$ supposed given). We sum up the results in

THEOREM III. *Let the functions $g_n(z)$ satisfy Conditions* A *in the circle $|z| \leqq R$. Then any function $f(z)$ analytic in $|z| < \rho \leqq R$ can be expanded in one and only one way by a series of the form*

$$f(z) = \sum_{n=0}^{\infty} c_n g_n(z)$$

convergent in $|z| < \rho$. Moreover, the coefficients c_n are determined by the formula

$$c_n = \frac{W[g_0(z), g_1(z), \cdots, g_{n-1}(z), f(z)]}{W[g_0(z), g_1(z), \cdots, g_{n-1}(z)]}\Bigg|_{z=0},$$

or by the recursion formula

$$c_n = f^{(n)}(0) - c_0 g_0^{(n)}(0) - c_1 g_1^{(n)}(0) - \cdots - c_{n-1} g_{n-1}^{(n)}(0).$$

It is scarcely necessary to point out that (3) reduces to a Taylor's series if $h_n(z) \equiv 0$, when Conditions A are surely satisfied. The existence of other functions $g_n(z)$ satisfying these conditions can not be held in question, so that (3) is a bona fide generalization of Taylor's series.

4. **Relation of the series to Birkhoff's series.** In order to make the proposed comparison we begin by stating

BIRKHOFF'S THEOREM.* *Let $v_0(z), v_1(z), \cdots$ be a sequence of functions each of which is analytic in a circle $|z| \leqq 1$. If the series*

$$|v_0(z) - 1| + |v_1(z) - z| + |v_2(z) - z^2| + \cdots \tag{11}$$

converges uniformly for $|z| = 1$ to a value less than unity, every function $f(z)$ analytic in $|z| \leqq 1$ can be represented by an absolutely and uniformly convergent series of the form

$$c_0 v_0(z) + c_1 v_1(z) + \cdots. \tag{12}$$

* We have altered the notation in order to avoid confusion. Also we state the result for the unit circle rather than for a circle of arbitrary radius in order to facilitate the comparison.

Here c_n is given by

$$c_n = \frac{1}{2\pi i}\int_C \frac{g(t)}{t^{n+1}}dt$$

where the integration is taken over the circle C, $|t|=1$, in the positive sense, and where $g(z)$ is the solution of the integral equation

$$f(z) = g(z) + \frac{1}{2\pi i}\int_C \left(\sum_{n=0}^{\infty} \frac{v_n(z) - z^n}{t^{n+1}}\right) g(t)dt.$$

We wish to point out the resemblance of series (12) to series (3) by showing that under the conditions imposed on (11) $v_n(z)$ must have exactly n zeros in the circle $|z|<1$. For, since the sum of the series (11) is less than unity on the unit circle, the general term must surely be less than unity there,

$$|v_n(z) - z^n| < 1, \quad |z| = 1. \tag{13}$$

Let z trace the unit circle once in the counter-clockwise sense. The change in the argument of $v_n(z)$ thus produced multiplied by $1/(2\pi)$ gives the number of zeros of $v_n(z)$ inside the circle. Clearly

$$\text{arc } v_n(z) = \text{arc } z^n[1 + \psi(z)/z^n] = n \text{ arc } z + \text{arc } [1 + \psi(z)/z^n],$$

where

$$\psi(z) = v_n(z) - z^n.$$

By (13)

$$|\psi(z)/z^n| < 1, \qquad |z| = 1,$$

and hence the change in the argument of $[1+\psi(z)/z^n]$ must be zero as z traces the unit circle. The change in arc z^n is $2n\pi$, so that $v_n(z)$ must have just n zeros inside the unit circle. These zeros need not be coincident at the origin,* however, as is the case with $g_n(z)$. We shall show that for the case in which all the zeros of $v_n(z)$ *are* concentrated at the origin, Conditions A include cases not included in Birkhoff's Conditions. This may be done by the simple example $h_n(z)=z/(n+1)$. Then

$$|n!g_n(z) - z^n| = 1/(n+1), \qquad |z| = 1.$$

Hence series (11) diverges if $v_n(z)=n!g_n(z)$. But Conditions A are satisfied. The expansion of $f(z)$ in a series (3) is assured by Theorem III, whereas Birkhoff's theorem is not applicable.

* The example $v_n(z)=z^n-1/2^n$ shows this.

BOSTON UNIVERSITY
COLLEGE OF LIBERAL ARTS
LIBRARY

If we restrict the functions $h_n(z)$ further by the inequality

$$|h_n(z)| < 1/2^{n+1}, \quad |z| = 1, \quad h_n(0) = 0,$$

both theorems are applicable. Since the expansion of a function $f(z)$ in a series (3) was shown to be unique, one may obtain an interesting identity by comparing the two determinations of the coefficients. That is, if

$$g_n(z) = \frac{z^n}{n!}[1 + h_n(z)], \quad |h_n(z)| < 1/2^{n+1}, \quad |z| = 1, \quad h_n(0) = 0,$$

then

$$\frac{W[g_0(z), g_1(z), \cdots, g_{n-1}(z), f(z)]}{W[g_0(z), g_1(z), \cdots, g_{n-1}(z)]}\Bigg|_{z=0} = \frac{n!}{2\pi i}\int_C \frac{g(t)}{t^{n+1}}dt,$$

where $g(z)$ satisfies the integral equation

$$f(z) = g(z) + \frac{1}{2\pi i}\int_C \left[\sum_{n=0}^{\infty} \frac{n!g_n(z) - z^n}{t^{n+1}}\right] g(t)dt.$$

C is the circle $z=1$, and the integration is in the positive sense.

UNIVERSITY OF MINNESOTA,
MINNEAPOLIS, MINN.

NOTE ON THE EXPANSION OF ANALYTIC FUNCTIONS IN SERIES OF POLYNOMIALS AND IN SERIES OF OTHER ANALYTIC FUNCTIONS*

BY

J. L. WALSH

On several occasions† the present writer has studied the expansion of functions of a complex variable in series of more or less arbitrary functions. The results obtained are connected with Birkhoff's generalization of Taylor's series,‡ but yield additional results when interpreted, after conformal transformation of the plane, as the expansion of arbitrary functions in series of polynomials. The object of the present note is to prove the following theorem, which is easily established by the methods already used, but seems to go in some directions beyond results proved by other writers§ concerning *polynomials associated with a region bounded by an arbitrary Jordan curve*:

In the plane of the complex variable z let C be an arbitrary Jordan curve in whose interior the origin lies. Then there exists a set $\{p_k(z)\}$ of polynomials in z having respectively k-fold roots at the origin and no other roots in or on C, and a set of functions $\{s_k(z)\}$, each a polynomial in $1/z$ of degree $k+1$, and such that

$$\int_{C'} s_k(z)p_i(z)dz = \begin{cases} 0, & i \neq k, \\ 1, & i = k, \end{cases}$$

where the integral is taken in the positive sense over an arbitrary rectifiable Jordan curve C' interior to C but which contains the origin in its interior.

---

* Presented to the Society, September 7, 1928; received by the editors in July, 1928.

† These Transactions, vol. 22 (1921), pp. 230–239; vol. 26 (1924), pp. 155–170; vol. 30 (1928), pp. 307–332. We shall refer to these papers as I, II, III respectively.

‡ Paris Comptes Rendus, vol. 164 (1917), pp. 942–945. A number of later writers have studied the same subject, under somewhat different hypotheses from those of Birkhoff. Detailed references are given by Widder, in the present number of these Transactions, pp. 43–52.

§ References to Faber and Szegö are given in II; see also Fejér, *Interpolation und konforme Abbildung*, Göttinger Nachrichten, 1918, pp. 319–331, and Carleman, *Über die Approximation analytischer Funktionen*, Arkiv för Matematik, Astronomi och Fysik, vol. 17 (1922–23). Fejér considers the development of functions analytic in the *closed* regions, Carleman of functions analytic interior to the given regions, integrable and with an integrable square (in the sense of surface integral) over the closed regions.

For developments in functions more general than polynomials, see Bergmann, Mathematische Annalen, vol. 86 (1922), pp. 238–271; Bochner, Mathematische Zeitschrift, vol. 14 (1922), pp. 180–207.

If $F(z)$ is an arbitrary function analytic interior to C, then the series

$$(1) \qquad \sum_{k=0}^{\infty} a_k p_k(z), \quad a_k = \int_{C'} F(z) s_k(z) dz,$$

converges to the sum $F(z)$ throughout the interior of C; the convergence is uniform in any closed region interior to C. The coefficients a_k in (1) can also be written in the form

$$a_k = A_0^{(k)} F(0) + A_1^{(k)} F'(0) + \cdots + A_k^{(k)} F^{(k)}(0),$$

where $A_i^{(k)}$ is a constant independent of $F(z)$, and where $F^{(i)}(0)$ indicates the ith derivative of $F(z)$ at the origin.

The interior of C can be mapped one-to-one and conformally on the interior of the unit circle γ in the x-plane by some transformation $x=\phi(z)$, $z=\psi(x)$, where $\phi(0)=0$, and the transformation is continuous in the closed regions involved. In general we denote by C_ρ the transform of the circle $|x|=\rho$, where $0<\rho<1$.

If $F(z)$ is an arbitrary function analytic interior to C_ρ, continuous in the corresponding closed region, and satisfies a Lipschitz condition on C_ρ, then the series (1), where C' is interior to C_ρ, converges uniformly in this closed region to the value $F(z)$. If $F(z)$ is required merely to be analytic interior to C_ρ, and is continuous in the corresponding closed region, then (1) converges uniformly, when summed by the method of Cesàro, to the value $F(z)$ in the closed region bounded by C_ρ. If $F(z)$ is subjected to no other condition than that of being analytic interior to C_ρ, series (1) converges interior to C_ρ to the sum $F(z)$; the convergence is uniform in any closed region interior to C_ρ.

*If $F(z)$ is an arbitrary function defined on C_ρ, integrable in the sense of Lebesgue, and if the condition**

$$\int_{C_\rho} F(z) z^k dz = 0 \qquad (k = 0,1,2,\cdots),$$

is satisfied, then the two series

$$(2) \qquad \sum_{k=0}^{\infty} a_k p_k[\psi(x)], \quad a_k = \int_{|x|=\rho} F[\psi(x)] s_k[\psi(x)] \psi'(x) dx,$$

$$\sum_{k=0}^{\infty} a_k' x^k, \qquad a_k' = \frac{1}{2\pi i} \int_{|x|=\rho} F[\psi(x)] x^{-k-1} dx,$$

---

* As in III, Theorem IX, no condition is here necessary if we set

$$a_k = \int_{|x|=\rho} F[\psi(x)] P_k(x) dx$$

instead of (2).

have essentially the same convergence properties on and within the circle $|x|=\rho$, *in the sense that in this closed region their term-by-term difference converges absolutely and uniformly to the sum zero.*

An arbitrary series of the form

$$\sum_{k=0}^{\infty} g_k p_k(z) \tag{3}$$

which converges for a single point z *on* C_ρ, *converges uniformly interior to* $C_{\rho'}$, *if* $\rho'<\rho$. *If* (3) *diverges for a point* z *on* C_ρ, *that series diverges for all points* z *exterior to* C_ρ *and on or interior to* C. *If in general we set*

$$\limsup_{k\to\infty} |g_k|^{1/k} = \frac{1}{\rho},$$

then if $\rho\leqq 1$, *series* (3) *converges for* z *interior to* C_ρ *and diverges for* z *exterior to* C_ρ *but on or interior to* C; *if* $\rho>1$, *series* (3) *converges for* z *on or interior to* C. *If* $0<\rho<1$, *some singular point of the function represented by the series lies on the curve* C_ρ.

If (3) *converges for the value* $z=z_1$ *on* C_ρ, *then that series converges uniformly in the closed region bounded by two arbitrary line segments whose interiors are interior to* C_ρ *and which terminate in* z_1, *and by an arc of a curve* $C_{\rho'}$, *where* $\rho'<\rho$. *If* (3) *converges uniformly on an arc* z_1z_2 *of the curve* C_ρ, *then that series converges uniformly in the closed region bounded by that arc, by two arbitrary line segments whose interiors are interior to* C_ρ *and which terminate respectively in* z_1 *and* z_2, *and by an arc of the curve* $C_{\rho'}$, *where* $\rho'<\rho$.

If (3) *is such that* $\lim_{k\to\infty} kg_k\rho^k=0$, *and if for approach along the normal* to* C_ρ *to the point* z_1 *on* C_ρ *we have*

$$\lim_{z\to z_1} f(z) = g,$$

where $f(z)$ *denotes the value of the series* (3), *which is convergent, under the hypothesis, for* z *interior to* C_ρ, *then we have also*

$$\sum_{k=0}^{\infty} g_k p_k(z_1) = g.$$

An arbitrary series (3), *convergent for a single value of* z *interior to* C *but not the origin, is the unique expansion of form* (1) *of some function* $F(z)$ *analytic on and within some curve* C_ρ.

---

* A more general theorem might easily be stated; see III, p. 313, third footnote. In III, Theorem VI, for g_k/k read kg_k, and in III, Theorem IX, for $g_k(k\rho^k)^{-1}$ read $kg_k\rho^k$.

Almost this entire theorem results from the two facts that the functions $\phi(z)$ and $\psi(x)$ are continuous in the closed interiors of C and γ respectively, and that in the former closed region $\phi(z)$ or any power of it can therefore be uniformly approximated as closely as desired by means of a polynomial in z; this approximating polynomial can even be chosen so that it has a zero at the origin whose multiplicity is the power of $\phi(z)$. For the function $[\phi(z)/z]^k$ is analytic interior to C, continuous in the corresponding closed region, hence in that region can be approximated as closely as desired by a polynomial in z. This gives directly the polynomial $p_k(z)$ (as in II, p. 163, and III, p. 316) which has a k-fold root at the origin and is such that

$$|\, p_k(z) - [\phi(z)]^k \,| \leqq \epsilon_k \tag{4}$$

is satisfied for all z on or interior to C. Inequality (4) implies, by virtue of III, §6, that the polynomials $p_k(z)$ just considered satisfy the requirements of III, Theorem IX, where an *arbitrary* curve C_ρ of the present theorem plays the rôle of the curve C of III. The theorem as stated follows, then, with the exception of a minor point about to be mentioned, because the same set of polynomials $p_k(z)$ can be considered independent of ρ for any curve C_ρ.

In II and III there were considered as given functions for expansion not arbitrary functions integrable in the sense of Lebesgue, but merely functions integrable and with an integrable square. The more general functions might easily have been treated there, however, as we shall now see by reference to I.

We assume the series $\sum \epsilon_n$ to be dominated by a convergent geometric series, $\epsilon_n \leqq R\tau^n$, $\tau<1$, so we have by III, p. 309, inequality (8),

$$|\, c_{nk} \,| \leqq \frac{R\tau^n}{(1+\epsilon)^k}.$$

In the notation of I,* we have also (I, p. 234, (13) and (11))

$$|\, c_{ki} + \bar{d}_{ik} \,| \leqq \frac{\gamma_i p_k}{1-p}, \quad \text{where} \quad \gamma_i^2 = \sum_{j=0}^{\infty} |\, c_{ji} \,|^2, \quad p_k^2 = \sum_{j=0}^{\infty} |\, c_{kj} \,|^2.$$

Then the two double series $\sum_{n,k=0}^{\infty} |\, c_{nk} \,|$, $\sum_{i,k=0}^{\infty} |\, d_{ik} \,|$ are absolutely convergent. The formal work of I, pp. 235–236, can now be justified immediately. If (in the notation of I) the function $f(x)$ is integrable in the sense of Lebesgue, the coefficients a_k are bounded in their totality; in fact a well known theorem due to Lebesgue asserts that $\lim_{k\to\infty} a_k = 0$. The convergence of the

* Slightly altered, for in I we consider only real functions, but in II and III complex functions as well. Compare II, p. 160.

The present inequalities replace those of III, p. 309, (10), (11), in which d_{ki} should be replaced by $\bar{d}_{ki}$, and p_k by γ_k.

series in the right-hand member of (16) (I, p. 236) follows now at once, for a dominant convergent series of positive terms exists.

This proves, then, that *in* III, *Theorem* I, *the two series* (3) *and* (4) *have for* $|x| \leqq 1$ *essentially the same convergence properties, provided merely that* $F(x)$ *is integrable on* γ *in the sense of Lebesgue.* The theorem of the present paper contains an application of this result.

The present theorem suggests, and in fact is easily proved by means of a simple modification of III, Theorems I–VIII. This modification, which we do not present in detail, replaces the requirement of III, Theorem I, that the functions $p_k(x)$ should be analytic for $|x| \leqq 1+\epsilon$, by the requirement that these functions should be continuous for $|x| \leqq 1$, analytic for $|x| < 1$. Expansions in terms of the functions $p_k(x)$ are then studied only for $|x| \leqq 1$, and the equivalence of expansions of arbitrary functions is considered, so far as the present methods apply without modification, only on circles $|x| = \rho < 1$.

In the theorem of the present note, the convergence of series (1) and (3) *on the boundary of* C deserves special study, with particular reference to the nature of the curve C (rectifiability, existence of tangents, etc.), although if C itself is analytic, a fairly complete study is of course contained in III. We make one further elementary remark in this connection. If C possesses forward and backward tangents at a point z_1, the transformation $x = \phi(z)$ is quasi-conformal at that point.* It follows, then, from the proof given in III, pp. 311, 312, that *if* (3) (notation of the present paper) *converges for a point* z_1 *of* C, *and if* C *possesses forward and backward tangents at* z_1, *then* (3) *converges uniformly in the closed region bounded by two arbitrary line segments interior to* C *but not tangent to* C *terminating in* z_1, *and by an arc of an arbitrary curve* C_ρ, $\rho < 1$.

Similarly, if (3) converges uniformly on an arc of C bounded by z_1 and z_2, and if C possesses forward and backward tangents at z_1 and z_2, then (3) converges uniformly in the closed region bounded by two arbitrary line segments interior to C but not tangent to C, terminating in z_1 and z_2 respectively, and by an arc of an arbitrary curve C_ρ.

---

* Carathéodory, Schwarz Festschrift, 1914, pp. 19–41, §20.

HARVARD UNIVERSITY
CAMBRIDGE, MASS.

A GENERAL PROBLEM OF MINIMIZING AN INTEGRAL WITH DISCONTINUOUS INTEGRAND*

BY

C. F. ROOS†

1. Introduction. There are several well known minimum problems which do not readily yield to the classical analysis of the calculus of variations. In some of these, such as the brachistochrone problem of determining the curve of quickest descent from a fixed curve to a fixed point, the integrand of the integral to be minimized is a function of an end value which is variable.‡ In others, such as that of finding a point π and a curve through this point and two fixed points such that the area of the surface of revolution formed by rotating this curve and the ordinate of π about the OX axis shall have as small a value as possible, the expression to be minimized is the sum of a point function and a definite integral.§ A more general theory which includes this latter problem as a special case has been given by E. H. Clarke.‖ In order to develop a general dynamical theory of depreciation it is desirable to consider a very general problem which somewhat resembles each of the above types, but which does not yield to the analysis of either.¶

The depreciation problem as well as the others referred to above can all be considered as special cases of a very general problem with discontinuous integrand, which also includes the most general problems of both the Lagrange and Mayer type with general boundary conditions.** In this paper it is proposed to develop a theory for minimizing an integral whose integrand is a discontinuous function sufficiently general in nature to include all of the the above mentioned types of problems and many others.††

* Presented to the Society, April 16, 1927; received by the editors in December, 1927.

† National Research Fellow.

‡ Bliss, *Calculus of Variations*, pp. 78–79.

§ Mary E. Sinclair, *Concerning a compound discontinuous solution in the problem of the surface of revolution of minimum area*, Annals of Mathematics, (2), vol. 10, p. 55.

‖ E. H. Clarke, *On the minimum of the sum of a definite integral and a function of a point*, Doctoral Dissertation, University of Chicago.

¶ C. F. Roos, *A mathematical theory of depreciation and replacement*, American Journal of Mathematics, January, 1928. See also Roos, *The problem of depreciation in the calculus of variations*, Bulletin of the American Mathematical Society, March–April, 1928.

** O. Bolza, *Über den "Anormalen Fall" beim Lagrangeschen und Mayerschen Problem mit gemischten Bedingungen und variablen Endpunkten*, Mathematische Annalen, vol. 74 (1913), pp. 430–446.

†† For the first discussion of a problem with discontinuous integrand see Mason and Bliss, *A problem of the calculus of variations in which the integrand is discontinuous*, these Transactions, vol. 7 (1906), pp. 325–336.

The expression to be minimized is the sum of two integrals taken from x_1 to a variable limit ω, and from ω to x_2, respectively. It may also be regarded as a single integral with an integrand which is discontinuous at the point ω, $y(\omega)$ of the minimizing curve. The integrand in addition to containing the usual arguments x, y and dy/dx contains also the end values x_1, $y(x_1)$, x_2, $y(x_2)$ and the corner values ω, $y(\omega)$.

In the work which follows we will find it convenient to use the notation of the tensor calculus to indicate summations. It will be understood, therefore, that when a subscript occurs twice in a term that term is to be summed for all values of the subscript.

2. **Statement of the problem.** It is desired to find among the values ω and arcs $y_i=y_i(x)$ $(i=1, \cdots, n;\ x_1 \leqq x \leqq x_2)$, satisfying m differential equations,

$$\phi_\alpha(x,\ y,\ y') = 0 \qquad (\alpha = 1, \cdots, m) \tag{1}$$

in x and y, and having end points satisfying end equations

$$\theta_\mu(\rho_1, \gamma_1, \rho_2, \gamma_2) = 0 \qquad (\mu = 1, \cdots, \rho \leqq 2n+2), \tag{2}$$

one which minimizes an expression of the form

$$I = \int_{x_1}^{\omega} g(x, y, y', \rho_1, \gamma_1, \rho_2, \gamma_2, \omega, y(\omega))dx + \int_{\omega}^{x_2} h(x, y, y', \rho_1, \gamma_1, \rho_2, \gamma_2, \omega, y(\omega))dx, \tag{3}$$

where for convenience in notation the set $(y_1, \cdots, y_n, y_1', \cdots, y_n')$ has been represented by (y, y'); γ_1 and γ_2 are n-partite numbers which for a particular curve $y_i(x)$ stand respectively for the sets $y_1(x_1), \cdots, y_n(x_1)$, and $y_1(x_2), \cdots, y_n(x_2)$, and, finally, primes denote derivatives with respect to x.

3. **Admissible arcs and variations.** In the analysis which follows we will need the following further hypotheses:*

(a) the functions $y_i(x)$ defining the minimizing arc E_{12} are continuous on the interval x_1, x_2, and this interval can be subdivided into a finite number of parts on each of which the functions have continuous derivatives;

(b) in a neighborhood R of the values x, y, y', ω, $y(\omega)$, x_1, $y(x_1)$, x_2, $y(x_2)$ on the minimizing arc the functions g, h and ϕ_α have continuous derivatives up to and including those of the second order;

---

* Bliss, *Lectures on the Calculus of Variations*, University of Chicago, summer quarter, 1925, mimeographed by O. E. Brown, Northwestern University, Evanston, Ill.; this will be cited as Bliss, *Lectures*.

(c) at every element (x, y, y') on E_{12} the $(m\times n)$-dimensional matrix $\partial\phi_\alpha/\partial y_i'$ has rank m, and

(d) the functions θ_μ have continuous derivatives up to and including those of the second order near the end values $(x_1, y(x_1), x_2, y(x_2))$ of the minimizing arc, and at these end values the $p\times(2n+2)$-dimensional matrix

$$\| \partial\theta_\mu/\partial x_1 \quad \partial\theta_\mu/\partial y(x_1) \quad \partial\theta_\mu/\partial x_2 \quad \partial\theta_\mu/\partial y(x_2) \|$$

has rank p.

An arc $y_i=y_i(x)$, $i=1, \cdots, n$; $x_1\leqq x\leqq x_2$, will be called an *admissible arc* if it is given by functions $y_i(x)$, defined on an interval (x_1, x_2), and by a number ω_0 such that $x_1<\omega_0<x_2$ and such that the $y_i'(x)$ are continuous on the intervals (x_1, ω_0) and (ω_0, x_2).

Let us suppose that we have a one-parameter family of admissible arcs given by functions

$$y_i(x, b), \quad x_1(b), \quad x_2(b) \quad \text{and} \quad \omega(b) \tag{4}$$

defined on an interval (b_1, b_2) such that $y_i(x, b_0)=y_i(x)$, $x_1(b_0)=x_1$, $x_2(b_0)=x_2$ and $\omega(b_0)=\omega_0$, and such that the functions $y_i(x, b)$ have continuous first-order derivatives with respect to x on the intervals $[x_1(b), \omega(b)]$ and $[\omega(b), x_2(b)]$. We define the functions $\eta_i(x)=\partial y_i(x, b_0)/\partial b$, $\xi_1=x_1'(b_0)$, $\xi_2=x_2'(b_0)$, $\xi_3=\omega'(b_0)$ as the *variations of the family along* E_{12}.

Let us denote by $\Omega_i(x)$ the variation $\eta_i(x)$ of $y_i(x, b)$ for the interval $x_1\leqq x\leqq\omega$ and by $\pi_i(x)$ the corresponding variation of $y_i(x, b)$ for the interval $\omega\leqq x\leqq x_2$, and by (u, u') the set (y, y') for $x_1\leqq x\leqq\omega$ and by (v, v') the corresponding set (y, y') for $\omega\leqq x\leqq x_2$. We do not assume that $\Omega_i(\omega)=\pi_i(\omega)$.

The *equations of variation for the functions* ϕ_α are defined by

$$\psi_\alpha = (\partial\phi_\alpha/\partial y_i)\eta_i + (\partial\phi_\alpha/\partial y_i')\eta_i' = 0 \qquad (\alpha = 1, \cdots, m) \tag{5}$$

where it is understood that $\eta_i=\Omega_i$ for $x_1\leqq x\leqq\omega$ and $\eta_i=\pi_i$ for $\omega\leqq x\leqq x_2$, and the coefficients $\partial\phi_\alpha/\partial y_i$, $\partial\phi_\alpha/\partial y_i'$ have as arguments the functions $y_i(x)$ defining the minimizing arc E_{12}.

The equations of variation for the functions θ_μ are

$$\begin{aligned}[\partial\theta_\mu/\partial x_1 + (\partial\theta_\mu/\partial y_i(x_1))y_i'(x_1)]\xi_1 + [\partial\theta_\mu/\partial y_i(x_1)]\eta_i(x_1)\\ + [\partial\theta_\mu/\partial x_2 + (\partial\theta_\mu/\partial y_i(x_2))y_i'(x_2)]\xi_2 + [\partial\theta_\mu/\partial y_i(x_2)]\eta_i(x_2) = 0.\end{aligned} \tag{6}$$

A set of functions $\eta_i(x)$ with the continuity properties described in (a) except at $x=\omega_0$, and satisfying the above equations of variation, (5) and (6), we define as an *admissible set of variations*.

THEOREM 1. *For every set of admissible variations $\eta_i(x)$ along the arc E_{12} there exists a one-parameter family,* (4), *of admissible arcs containing E_{12} for the value $b=0$ and having the functions $\eta_i(x)$ as its variations along E_{12}. For this family the functions $y_i(x, b)$ are continuous and have continuous derivatives with respect to b for all values, (x, b), near those defining E_{12}, and the derivatives $y_{ix}(x, b)$ have the same property except at the values of x on the curve defined by $x=\omega(b)$, $y_i=y_i(\omega(b), b)$ and, possibly, at the values of x defining other corners of E_{12}.*

This theorem can be proved by slightly modifying a proof given by Bliss.* Following his example we enlarge the system $\phi_\alpha=0$ to have the form

$$\phi_1 = 0, \cdots, \phi_m = 0, \quad \phi_{m+1} = z_{m+1}, \cdots, \phi_n = z_n, \tag{7}$$

where $z_{m+1}, \cdots, z_n$ are new variables, and $\phi_{m+1}, \cdots, \phi_n$ are new functions of (x, y, y') such that the functional determinant $|\partial\phi_i/\partial y_k'|$ is different from zero along E_{12}.† By means of the last $n-m$ of these equations the functions $y_i(x)$ belonging to E_{12} define a set of functions $z_r(x)$ $(r=m+1, \cdots, n)$.

We have a corresponding system of equations of variation,

$$\psi_1 = 0, \cdots, \psi_m = 0, \quad \psi_{m+1} = \zeta_{m+1}, \cdots, \psi_n = \zeta_n, \tag{8}$$

along E_{12}. The last $n-m$ of these define a set $\zeta_r(x)$ corresponding to every set of admissible variations $\eta_i(x)$.

Since $|\partial\phi_i/\partial y_k'| \neq 0$ along E_{12}, the existence theorems for differential equations tell us that the system‡

$$\phi_\alpha = 0, \quad \phi_r = z_r(x) + b\zeta_r(x) \tag{9}$$

determines uniquely a one-parameter family of solutions,

$$y_i = U_i(x, b) \qquad (x_1(b) \leqq x \leqq \omega(b)), \tag{10}$$

with initial conditions

$$y_i(x_1) + b\eta_i(x_1) = U_i(x_1, b),$$

---

* Bliss, *Lectures*, pp. 4–6.

† For a proof of the possibility of this adjunction see Bliss, *The problem of Mayer with variable end points*, these Transactions, vol. 19 (1918), p. 312.

‡ Bolza, *Vorlesungen über Variationsrechnung*, pp. 168 ff.

Bliss, Annals of Mathematics, (2), vol. 6 (1904), p. 49.

Bliss, Bulletin of the American Mathematical Society, vol. 25 (1918), p. 15.

where $x=\omega(b)$ is the value of x along a curve of corners $x=\omega(b)$, $u_i=v_i=y_i(\omega(b), b)$ with $\omega(b)>x_1(b)$ and $\omega(b_0)=\omega_0$ but $\omega(b)$ otherwise arbitrary.

For the interval $(\omega(b)\leqq x\leqq x_2(b))$ there is a unique one-parameter family of solutions $y_i=V_i(x, b)$ having initial conditions at $x=\omega$,

$$V_i(\omega(b),b) = U_i(\omega(b),b). \tag{11}$$

This family consisting of $U_i(x, b)$ on $(x_1\leqq x\leqq\omega)$ and $V_i(x, b)$ on $(\omega\leqq x\leqq x_2)$ contains E_{12} for $b=0$ and has variations which satisfy the equations (8) with the functions $\zeta_r(x)$. The variations of the family are, therefore, identical with the functions $\eta_i(x)$ originally prescribed, since when the ζ_r are given, there is only one set of solutions of equations (8) with given initial values $\eta_i(x_1)$ at $x=x_1$.

COROLLARY. *For a matrix*

$$\left\| \begin{matrix} \xi_1^{(1)} & \cdots & \xi_1^{(p+1)} \\ \xi_2^{(1)} & \cdots & \xi_2^{(p+1)} \\ \eta_1^{(1)} & \cdots & \eta_1^{(p+1)} \\ \cdots & \cdots & \cdots \\ \eta_n^{(1)} & \cdots & \eta_n^{(p+1)} \end{matrix} \right\|,$$

whose columns are sets of admissible variations, there exists a $(p+1)$-parameter family, $y_i=y_i(x, b_1, \cdots, b_{p+1})$, $x_1(b_1, \cdots, b_{p+1})\leqq x\leqq x_2(b_1, \cdots, b_{p+1})$, containing E_{12} for $(b_1, \cdots, b_{p+1})=(0, \cdots, 0)$ and having the sets $\xi_1^{(\omega)}$, $\xi_2^{(\omega)}$ and $\eta_i^{(\omega)}(x)$ as its variations along E_{12} with respect to the parameter b_ω. The continuity properties are similar to those described in the above theorem.

This corollary is proved with the equations $\phi_\alpha=0, \cdots, \phi_r=z_r+b_1\zeta_r^{(1)}+\cdots+b_{p+1}\zeta_r^{(p+1)}$ replacing equations (9).

4. **First variation of the integral I.** If the one-parameter family of admissible arcs (4) be substitued in (3), we obtain for the first variation of the resulting expression for $b=b_0=0$

$$\begin{aligned} I_1(\xi,\eta) = {} & \int_{x_1}^{\omega_0}[(\partial g/\partial u_i)\Omega_i + (\partial g/\partial u_i')\Omega_i']dx + \int_{\omega_0}^{x_2}[(\partial h/\partial v_i)\pi_i \\ & + (\partial h/\partial v_i')\pi_i']dx + S_{1i}(g,h)\Omega_i(x_1) + S_{2i}(g,h)\pi_i(x_2) + L_1(g,h)\xi_1 \\ & + L_2(g,h)\xi_2 + M(g,h)\omega'(0) + N_i(g,h)\frac{dy_i(\omega(0),0)}{db}, \end{aligned} \tag{12}$$

where as notation

$$S_\sigma(g,h) = \int_{x_1}^{\omega_0}(\partial g/\partial y_i(x_\sigma))dx + \int_{\omega_0}^{x_2}(\partial h/\partial y_i(x_\sigma))dx \qquad (\sigma = 1,2);$$

$$L_\sigma(g,h) = (\sigma-2)g(x_1) + (\sigma-1)h(x_2) + \int_{x_1}^{\omega_0}[\partial g/\partial x_\sigma + (\partial g/\partial y_i(x_\sigma))y_i'(x_\sigma)]dx + \int_{\omega_0}^{x_2}[\partial h/\partial x_\sigma + (\partial h/\partial y_i(x_\sigma))y_i'(x_\sigma)]dx;$$

$$M(g,h) = g(\omega_0) - h(\omega_0) + \int_{x_1}^{\omega_0}(\partial g/\partial\omega)dx + \int_{\omega_0}^{x_2}(\partial h/\partial\omega)dx;$$

$$N_i(g,h) = \int_{x_1}^{\omega_0}[\partial g/\partial u_i(\omega)]dx + \int_{\omega_0}^{x_2}[\partial h/\partial v_i(\omega)]dx;$$

$g(x_1)$ and $g(\omega_0)$ are the values of g at the points (1) and (ω_0) on E_{12}; $h(\omega_0)$ and $h(x_2)$ are the values of h at the points (ω_0) and (2) on E_{12}; i is an umbral index with range $1, \cdots, n$ when it occurs twice as a subscript, and σ is not umbral in L_σ.

Let us multiply each of the equations (8) by a function $\lambda(x)$; form the sum $\lambda_\alpha\psi_\alpha+\lambda_r\psi_r-\lambda_r\zeta_r$ where α and r are umbral indices with ranges $1, \cdots, m$ and $m+1, \cdots, n$, respectively, and then integrate this sum from x_1 to x_2. We obtain

$$\int_{x_1}^{\omega_0}[(\partial R/\partial u_i)\Omega_i + (\partial R/\partial u_i')\Omega_i']dx + \int_{\omega_0}^{x_2}[(\partial T/\partial v_i)\pi_i + (\partial T/\partial v_i')\pi_i']dx - \int_{x_1}^{x_2}\lambda_r\zeta_r dx = 0, \tag{13}$$

where, by definition, $R=\lambda_i\phi_i$, $i=1, \cdots, n$, on $x_1 \leqq x \leqq \omega$, and $T=\lambda_i\phi_i$ on $\omega \leqq x \leqq x_2$. It seems desirable to split the integral into two parts in this manner because this was done with the original integral I.

By (6) the equations of variation on E_{12} for the functions θ_μ are

$$K_\mu(\xi,\eta) = [(\partial\theta_\mu/\partial x_1) + (\partial\theta_\mu/\partial y_i(x_1))y_i'(x_1)]\xi_1 + [\partial\theta_\mu/\partial y_i(x_1)]\Omega_i(x_1) + [\partial\theta_\mu/\partial x_2 + (\partial\theta_\mu/\partial y_i(x_2))y_i'(x_2)]\xi_2 + [\partial\theta_\mu/\partial y_i(x_2)]\pi_i(x_2) = 0. \tag{14}$$

If a $(p+1)$-parameter family of admissible arcs $y_i=y_i(x, b_1, \cdots, b_{p+1})$, $x_1(b_1, \cdots, b_{p+1})$, $x_2(b_1, \cdots, b_{p+1})$, $\omega(b_1, \cdots, b_{p+1})$, containing E_{12} for $(b_1, \cdots, b_{p+1})=(0, \cdots, 0)$, be substituted in I and the functions θ_μ, these become functions of $b_1, \cdots, b_{p+1}$. If I_0 is the desired minimum yielded by E_{12}, then the first members of the equations

$$I(b_1, \cdots, b_{p+1}) = I_0 + q,$$
$$\theta_\mu(b_1, \cdots, b_{p+1}) = 0,$$

where q is a positive number, must have their functional determinant equal to zero for $(b_1, \cdots, b_{p+1}) = (0, \cdots, 0)$, since, otherwise, these equations would have solutions near this set of values for both positive and negative values of q. This functional determinant of rank $r < p+1$ is

$$(15) \qquad \begin{vmatrix} I_1(\xi^{(1)}, \eta^{(1)}) & \cdots & I_1(\xi^{(p+1)}, \eta^{(p+1)}) \\ K_1(\xi^{(1)}, \eta^{(1)}) & \cdots & K_1(\xi^{(p+1)}, \eta^{(p+1)}) \\ \cdots & \cdots & \cdots \\ K_p(\xi^{(1)}, \eta^{(1)}) & \cdots & K_p(\xi^{(p+1)}, \eta^{(p+1)}) \end{vmatrix},$$

and from its vanishing we argue that there exists a set of constants $d_1, \cdots, d_p, \lambda_0$, not all zero, such that the equation

$$(16) \qquad \lambda_0 I_1(\xi, \eta) + d_\mu K_\mu(\xi, \eta) = 0 \qquad (\mu \text{ umbral with range } 1, \cdots, p)$$

holds for every set of admissible variations $\xi_1, \xi_2, \omega'(0), \eta(x)$, since, otherwise, one of the columns of this determinant could be replaced by another making it of rank $r+1$.*

As notation let ν be an umbral index with range 1, 2. Let us define two new functions G and H by the respective equations

$$G(x, y, y', x_1, \cdots, y(x_2), \lambda) = \lambda_0 g + \lambda_i \phi_i,$$
$$H(x, y, y', x_1, \cdots, y(x_2), \lambda) = \lambda_0 h + \lambda_i \phi_i.$$

Since the variations (ξ, η) satisfy (13), the value of (16) is not altered if we add (13) to it. By the help of (12) we can, therefore, write

$$(17) \qquad \begin{aligned} &\int_{x_1}^{\omega_0} [(\partial G/\partial u_i)\Omega_i + (\partial G/\partial u_i')\Omega_i' - \lambda_r \zeta_r] dx \\ &+ \int_{\omega_0}^{x_2} [(\partial H/\partial v_i)\pi_i + (\partial H/\partial v_i')\pi_i' - \lambda_r \zeta_r] dx + [S_{\nu i}(g, h) \\ &+ d_\mu(\partial \theta_\mu/\partial y_i(x_\nu))]\eta_i(x_\nu) + [L_\nu(g, h) + d_\mu(\partial \theta_\mu/\partial x_\nu)]\xi_\nu \\ &+ [(\partial \theta_\mu/\partial y_i(x_1))y_i'(x_1)]\xi_1 + [(\partial \theta_\mu/\partial y_i(x_2))y_i'(x_2)]\xi_2 + M(g, h)\omega'(0) \\ &+ N_i(g, h) dy_i(\omega(0), 0)/db = 0. \end{aligned}$$

So far the functions $\lambda_i(x)$ have been entirely arbitrary.

5. **First necessary conditions.** We now proceed to determine the $\lambda_i(x)$ so that the equations

---

* Bliss, *Lectures*, p. 20.

$$\partial G/\partial u_i' = \int_{x_1}^{x} (\partial G/\partial u_i)dx + C_{i1} \qquad (x_1 \leqq x \leqq \omega), \tag{18}$$

$$\partial H/\partial v_i' = \int_{x}^{x_2} (\partial H/\partial v_i)dx + C_{i2} \qquad (\omega \leqq x \leqq x_2) \tag{19}$$

are satisfied for an arbitrarily selected set of constants C_{i1}, C_{i2}. This is possible because of the assumption that $|\partial\phi_i/\partial y_k'|$ is different from zero along E_{12}. For the details of the proof of the possibility of this choice of the λ_i we refer the reader to the mimeographed lecture notes of G. A. Bliss,* since the proof which Bliss gives there requires only slight alteration. For this choice of the functions $\lambda_i(x)$ equation (17) becomes

$$\begin{aligned} &-\int_{x_1}^{\omega_0} \lambda_r\zeta_r dx - \int_{\omega_0}^{x_2} \lambda_r\zeta_r dx + [P_\nu(x_\nu) + d_\mu(\partial\theta_\mu/\partial y_i(x_\nu))]\eta_i(x_\nu) + [L(x_\nu) \\ &+ d_\mu(\partial\theta_\mu/\partial x_\nu)]\xi_\nu + [d_\mu(\partial\theta_\mu/\partial y_i(x_1))y_i'(x_1)]\xi_1 + [d_\mu(\partial\theta_\mu/\partial y_i(x_2))y_i'(x_2)]\xi_2 \\ &+ M(g,h)\omega'(0) + [\partial G/\partial u'(\omega_0)]\Omega_i(\omega_0) - [\partial H/\partial v_i'(\omega_0)]\pi_i(\omega_0) \\ &+ [N_i(g,h)]dy_i(\omega_0,0)/db = 0, \end{aligned} \tag{20}$$

where

$$P_i(x_\nu) = (\nu - 2)(\partial G/\partial u_i'(x_1) + (\nu - 1)(\partial H/\partial v_i'(x_2)) + S_{\nu i}(g,h) \qquad (\nu = 1,2),$$

and $L(x_\nu) = L_\nu(g, h)$; $\partial G/\partial u_i'(x_1)$ and $\partial G/\partial u_i'(\omega_0)$ denote derivatives with respect to u_i' at the point (1) and (ω_0), respectively, on E_{12}, and $\partial H/\partial v_i'(\omega_0)$ and $\partial H/\partial v_i'(x_2)$ denote derivatives of H with respect to v_i' at the points (ω_0) and (2), respectively, on E_{12}, and where we have chosen the constants C_{i1} and C_{i2} so that equations (18) and (19) are satisfied, i.e., $C_{i1} = \partial G/\partial u_i'(x_1)$ and $C_{i2} = \partial H/\partial v_i'(x_2)$.

Since the extremal E_{12} is continuous at $x = \omega_0$ the equations $y_i(\omega(b), b) = u_i(\omega(b), b) = v_i(\omega(b), b)$ must hold, and, hence, by a differentiation

$$dy_i(\omega_0,0)/db = u_i'(\omega_0)\omega'(0) + \Omega_i(\omega_0) = v_i'(\omega_0)\omega'(0) + \pi_i(\omega_0). \tag{21}$$

If we substitute the values of $\Omega_i(\omega_0)$ and $\pi_i(\omega_0)$ defined by (21) in (20), we obtain as a necessary condition for a minimum of the integral I

$$\begin{aligned} &-\int_{x_1}^{\omega_0} \lambda_r\zeta_r dx - \int_{\omega_0}^{x_2} \lambda_r\zeta_r dx + [P_i(x_\nu) + d_\mu(\partial\theta_\mu/\partial y_i(x_\nu))]\eta_i(x_\nu) + [L(x_\nu) \\ &+ d_\mu(\partial\theta_\mu/\partial x_\nu)]\xi_\nu + [d_\mu(\partial\theta_\mu/\partial y_i(x_1))y_i'(x_1)]\xi_1 + [d_\mu(\partial\theta_\mu/\partial y_i(x_2))y_i'(x_2)]\xi_2 \\ &+ [-u'(\omega_0)\partial G/\partial u_i'(\omega_0) + v_i'(\omega_0)\partial H/\partial v_i'(\omega_0) + M(g,h)]\omega'(0) \\ &+ [\partial G/\partial u_i'(\omega_0) - \partial H/\partial v_i'(\omega_0) + N_i(g,h)]dy_i(\omega_0,0)/db = 0. \end{aligned} \tag{22}$$

* Bliss, *Lectures*, p. 7.

Since $\omega'(0)$ and $\Omega_i(\omega_0)$ are arbitrary, equations (21) show us that $dy_i(\omega_0, 0)/db$ is arbitrary. We can, therefore, say that the expression (22) must vanish for every choice of $\eta_i(x_\nu)$, ξ_ν, $\omega'(0)$, $dy_i(\omega_0, 0)/db$ and ζ_r, $r=m+1, \cdots, n$. It follows then that all of the coefficients of these quantities must vanish and, hence, in particular, that $\lambda_{m+1} \equiv \cdots \equiv \lambda_n \equiv 0$. It follows further with the help of (12) that

$$
\begin{aligned}
&- u_i'(\omega_0)(\partial G/\partial u_i'(\omega_0)) + v_i'(\omega_0)(\partial H/\partial v_i'(\omega_0)) + G(\omega_0) - H(\omega_0) \\
&\quad + \int_{x_1}^{\omega_0} (\partial G/\partial \omega)dx + \int_{\omega_0}^{x_2} (\partial H/\partial \omega)dx = 0, \\
&\partial G/\partial u_i'(\omega_0) - \partial H/\partial v_i'(\omega_0) + \int_{x}^{\omega_0} (\partial G/\partial u_i(\omega_0))dx \\
&\quad + \int_{\omega_0}^{x_2} (\partial H/\partial v_i(\omega_0))dx = 0 \qquad (i = 1, \cdots, n),
\end{aligned} \tag{23}
$$

where as already stated u_i refers to the extremal arc y_i for the interval $x_1 \leqq x \leqq \omega_0$, and v_i refers to the same arc y_i but for the interval $\omega_0 \leqq x \leqq x_2$; further that in the matrix

$$
\left\| \begin{matrix} L(x_1) & P_i(x_1) & L(x_2) & P_i(x_2) \\ \theta_{\mu x_1} + \theta_{\mu y_{i1}} y_{i1}' & \theta_{\mu y_{i1}} & \theta_{\mu x_2} + \theta_{\mu y_{i2}} y_{i2}' & \theta_{\mu y_{i2}} \end{matrix} \right\|, \tag{24}
$$

where $\theta_{\mu x_1} + \theta_{\mu y_{i\nu}} y_{i\nu}'$ stands for $\partial\theta_\mu/\partial x_1 + (\partial\theta_\mu/\partial y_i(x_\nu))y_i'(x_\nu)$, every determinant of order $p+1$ vanishes, and finally that the equations (18) and (19) are satisfied. The condition (24) must hold since the multipliers $1, d_1, \cdots, d_p$ satisfy all the linear equations whose coefficients are columns of the matrix. The rank of this matrix is unchanged when one column is multiplied by a factor and added to another, so that we can state the following theorem.

THEOREM 2. *For every minimizing arc for this general problem there exist sets of constants C_{i1} and C_{i2}, $i=1, \cdots, n$, and functions*

$$
\begin{aligned}
G(x, y, y', x_1, y(x_1), \cdots, y(x_2), \lambda) &= \lambda_0 g + \lambda_\alpha \phi_\alpha \qquad (x_0 \leqq x \leqq \omega_0), \\
H(x, y, y', x_1, y(x_1), \cdots, y(x_2), \lambda) &= \lambda_0 h + \lambda_\alpha \phi_\alpha \qquad (\omega_0 \leqq x \leqq x_2),
\end{aligned}
$$

such that the equations

$$
\begin{aligned}
\partial G/\partial u_i' &= \int_x^x (\partial G/\partial u_i)dx + C_{i1} \qquad (x_1 \leqq x \leqq \omega), \\
\partial H/\partial v_i' &= \int_x^{x_2} (\partial H/\partial v_i)dx + C_{i2} \qquad (\omega \leqq x \leqq x_2)
\end{aligned}
$$

are satisfied at every point of E_{12}, $x_1 \leqq x \leqq x_2$. The constant λ_0 and the functions

$\lambda_\alpha(x)$, $\alpha=1, \cdots, m$, *are not all identically zero on* E_{12}, *and are continuous except possibly at values of* x *defining corners of* E_{12}. *At the point* ω, $y_i(\omega)$, *at which there exists a corner, the equations* (23), *which are the analogues of the Weierstrass-Erdmann corner conditions, must be satisfied. Furthermore, the end values of* E_{12} *must be such that all determinants of order* $p+1$ *of the matrix*

$$\left\| \begin{matrix} L(x_1) - y'_{i1}P_i(x_1) & P_i(x_1) & L(x_2) - y'_{i2}P_i(x_2) & P_i(x_2) \\ \theta_{\mu x_1} & \theta_{\mu y_{i1}} & \theta_{\mu x_2} & \theta_{\mu y_{i2}} \end{matrix} \right\| \tag{25}$$

vanish.

These last conditions are the *transversality conditions*. It is easily seen that if the functions g and h are independent of the end and corner variables they reduce to the corresponding conditions obtained by Bliss.*

The corner conditions as obtained in this paper differ from the corner conditions usually obtained because of the occurrence of the integrals involving derivatives with respect to corner variables, so that if g and h are independent of these variables, the integral terms drop out. For such problems as Miss Sinclair's and E. H. Clarke's these derivatives reduce to terms simple enough to be integrated by inspection.†

It will be interesting to examine a few special cases of the very general problem of this paper. In as much as the paper was suggested by the problem of depreciation it will be well to begin with a short examination of this important special case.

6. **Some special cases.** In order to determine the best time at which to replace a machine, which is in operation, by another machine whose operating expense is different we must maximize an expression which can be thrown in the general form (3). This problem of replacement in dynamical economics is that of finding among the arcs $y_i=y_i(x)$ satisfying a differential equation of demand, $D(x, y_1, y_1', y_2, y_2')=0$, where x is the time, y_1 is the rate of production at the time x, and y_2 is the price at the time x, and end conditions

$$\theta_\mu(x_1, y_1(x_1), y_2(x_1), x_2, y_1(x_2), y_2(x_2)) = 0 \qquad (\mu = 1, \cdots, p \leqq 6),$$

a set which maximizes an expression of the form

$$V = \int_{x_1}^{\omega} [y_1y_2 - Q_1(x, y_1, y_2, y_1', y_2')]e^{\beta(x_1-x)}dx$$
$$+ \int_{\omega}^{x_2} [y_1y_2 - Q_2(x, y_1, y_2, y_1', y_2')]e^{\beta(x_1-x)}dx$$
$$+ C_1e^{\beta(x_1-\omega)} + C_2e^{\beta(x_1-x_2)},$$

* Bliss, *The problem of Mayer with variable end points*, loc. cit.

† See fourth and fifth footnotes on page 58.

where x is the time; V is the value at $x=x_1$ of the composite machine, consisting of a machine operating for the period $x_1 \leqq x \leqq \omega$, and a replacement machine operating for the period $\omega \leqq x \leqq x_2$; y_1 is the number of units of output per unit time; y_2 is the price of one unit of output at the time x; Q_1 is the cost of producing y_1 units of output for the machine called the operating machine; Q_2 is the cost of producing y_1 units of output for the replacement machine, and exp. $(\beta(x_1-x))$ etc. are discount factors to allow for interest.

Since ω and x_1 are constant so far as integration with respect to x is concerned, we may write this expression in the form

$$V = \int_{x_1}^{\omega} \left[(y_1 y_2 - Q_1) e^{\beta(x_1 - x)} + \frac{C_1 e^{\beta(x_1-\omega)} + C_2 e^{\beta(x_1-x_2)}}{\omega - x_1} \right] dx + \int_{\omega}^{x_2} (y_1 y_2 - Q_2) e^{\beta(x_1-x)} dx.$$

When V is in this form, the replacement problem is easily seen to be a special case of the general problem of this paper. In as much as I have presented a mathematical theory of depreciation and replacement in another paper, there is no need to carry the problem further here.*

As already mentioned, problems which require that the sum of a point function and a definite integral be minimized are special cases of a general problem discussed by Clarke.† His problem is that of finding among the arcs $y_i=y_i(x)$ satisfying end conditions $x_1=\alpha_1$, $x_2=\alpha_2$, $y_i(x_1)=\beta_{i1}$, $y_i(x_2)=\beta_{i2}$ $(i=1, \cdots, n)$, a set which minimizes the sum of a definite integral and a function of a point,

$$J = \phi(\omega, y(\omega)) + \int_{x_1}^{x_2} f(x, y, y') dx.$$

Although Clarke does not discuss the Lagrange problem analogous to the case above which he studied, the analysis of my paper is sufficiently general to apply to his problem. We may write J in the form

$$J = \int_{x_1}^{\omega} \left[f(x, y, y') + \frac{\phi(\omega, y(\omega))}{\omega - x_1} \right] dx + \int_{\omega}^{x_2} f(x, y, y') dx,$$

and again obtain a special case. For Clarke's problem the corner conditions (23) become

---

* Roos, *A mathematical theory of depreciation and replacement*, loc. cit.

† Clarke, *On the minimum of the sum* etc., loc. cit.

$$\phi_\omega + f(\omega, y(\omega), u'(\omega)) - f(\omega, y(\omega), v'(\omega)) - u'(\omega) f_u'(\omega, y(\omega), u'(\omega)) + v'(\omega) f_v'(\omega, y(\omega), u'(\omega)) = 0,$$

$$\phi_{y_i(\omega)} + f_{u_i'}(\omega, y(\omega), u'(\omega)) - f_{v_i'}(\omega, y(\omega), v'(\omega)) = 0 \qquad (i = 1, \cdots, n),$$

and these are identical with the conditions obtained by Clarke, except for notation.

Let us next consider special cases for which the functions g and h are independent of the corner variables and, furthermore, for which $g \equiv h$, i.e., for which there is no corner at $x = \omega$.

An important example of this type of problem is the following proposed by Bolza.

Minimize the expression

$$U = \int_{x_0}^{x_1} H(y_1, \cdots, y_n, y_1', \cdots, y_n') dx + G(y_{10}, \cdots, y_{n0}, y_{11}, \cdots, y_{n1}),$$

when the admissible curves are defined by functions $y_i = y_i(x)$ $(x_0 \leqq x \leqq x_1)$ which satisfy differential equations $\phi_\alpha(y_i, y_i') = 0$ $(\alpha = 1, \cdots, m)$, while the end points, y_{i0}, y_{i1}, satisfy conditions $\theta_\mu(y_{i0}, y_{i1}) = 0$ $(\mu = 1, \cdots, p)$. Bolza admits so-called finite equations $\psi_k(y_1, \cdots, y_n) = 0$ $(k = 1, \cdots, r < n - m)$, in addition to the equations $\phi_\alpha = 0$, but, as Bliss has shown, this gives no additional generalization.* Bolza's problem is the special case for which ϕ_α and θ_μ of the general problem do not depend upon x explicitly and for which

$$g \equiv h \equiv \frac{G(y_{10}, \cdots, y_{n0}, y_{11}, \cdots, y_{n1})}{x_1 - x_0} + H(y_1, \cdots, y_n, y_1', \cdots, y_n').$$

The integrand is continuous at $x = \omega$ and the corner conditions vanish identically unless $y_i(x)$ has a corner there, in which case they reduce to the Weierstrass-Erdmann corner conditions.†

In as much as Bolza has shown that his problem is sufficiently general to include as special cases the most general Lagrange problem with general boundary conditions for $G \equiv 0$, and the most general Mayer problem with general boundary conditions for the case $H \equiv 0$, $G \equiv y_{10}$, it appears that the problem treated in this memoir is sufficiently general to include a very large class of the problems of the calculus of variations.

7. **Further generalizations.** In the replacement problem discussed in the last paragraph we supposed that there was only one machine producing

---

* Bliss, *The problem of Mayer with variable end points*, loc. cit.

† J. Hadamard, *Leçons sur le Calcul des Variations*, vol. 1, p. 188.

the commodity or service under consideration at any time x. If we suppose several machines to be simultaneously in operation producing the same article, there are two problems of interest. If all owners agree to cooperate in making the sum of their profits a maximum the problem is a special case of the one considered in this paper. On the other hand if the machines are competing, the problem is one for which it is required to obtain partial maxima of as many integrals as there are competing machines. I have already considered the competition case for the special case in which g and h are independent of variable corner and end values.*

In any actual case we do not have all machines in operation for the same period of time, for new competitors are constantly entering and others are dropping out. Such a problem becomes difficult even for the cooperation case because we must maximize an expression which is the sum of say n integrals whose limits may or may not be the same and whose integrands depend upon different numbers of variables x and y_i. Each time a newcomer produces, the demand equation contains an additional variable, e.g., this producer's rate of production.

---

* Roos, *Generalized Lagrange problems in the calculus of variations*, these Transactions, April, 1928. See also Roos, *A mathematical theory of competition*, American Journal of Mathematics, vol. 47 (1925), pp. 163–175, and Roos, *A dynamical theory of economics*, Journal of Political Economy, vol. 35 (1927), pp. 632–656.

UNIVERSITY OF CHICAGO,
CHICAGO, ILL.

AN INTRODUCTION TO THE THEORY OF IDEALS IN LINEAR ASSOCIATIVE ALGEBRAS*

BY

C. C. MACDUFFEE

1. Introduction. With the development of the number theory of linear algebras, it was natural that attempts should be made to extend to these domains of integrity the theory of ideal numbers. But it is evident from the few domains which have been examined that one cannot expect an extension of this theory in its entirety. For instance, Hurwitz† has investigated the number theory of quaternions by using right and left ideals, and has found that they are powerless to introduce unique factorization into this algebra. Moreover, Speiser‡ has recently investigated the properties of right, left and two-sided ideals in semi-simple algebras, remarking in the introduction to his paper that some of the most remarkable properties of ideals are "but foreign adjuncts which are essentially restricted to algebraic number fields."

Although it is historically true that ideals were introduced into algebraic number theory to establish unique factorization, it should be observed that this is a secondary function of ideals. Primarily they establish the property that every two numbers have a greatest common divisor expressible linearly in terms of the numbers. In algebraic fields this property implies unique factorization but in the general linear algebra it does not—hence the success of the ideal theory in algebraic fields and its partial failure in the more general domain.

The method which is here used in developing the theory of ideals is different from the usual one. It depends upon a correspondence§ between ideals and matrices whose elements are rational integers, and the only kind of multiplication which is employed is ordinary matric multiplication. Ideal multiplication, which plays such an important rôle in the usual treatment of ideals in algebraic fields, but which causes so much difficulty in non-commutative domains, is not employed in this paper.

---

* Presented to the Society, September 9, 1926; received by the editors in September, 1927.

† A. Hurwitz, *Vorlesungen über die Zahlentheorie der Quaternionen*, Berlin, Springer, 1919.

‡ A. Speiser, Vierteljahrsschrift der Naturforschenden Gesellschaft in Zürich, vol. 71 (1926), pp. 8–49.

§ Applied to quadratic ideals in a recent paper by the author, Annals of Mathematics, (2), vol. 29 (1927–28), pp. 199–214.

The fundamental theorem in the present development (Theorem 4) serves as a substitute for the commutative law in the multiplication of these ideal matrices. From this starting point we proceed to establish for general algebras many of the properties which ideals in algebraic fields are known to possess, including the important concept of class number. Furthermore, we obtain some properties of ideals in all semi-simple algebras which have not previously been found even in the special case of algebraic fields.

The writer is indebted to his colleague Professor Blumberg for helpful suggestions.

2. Definition of ideal. We suppose that $\mathfrak{A}$ is a rational semi-simple algebra of order n with a principal unit, and that $\mathfrak{S}$ is a set of integral elements of order n in $\mathfrak{A}$ according to the definition of Dickson.* We suppose that the basal numbers $e_1, e_2, \cdots, e_n$ of $\mathfrak{A}$ form a basis for $\mathfrak{S}$, e_1 being the principal unit.† We shall define a left (right) ideal $\mathfrak{R}$ to be a set of numbers of $\mathfrak{S}$ which is closed under addition and subtraction and under multiplication on the left (right) by the numbers of $\mathfrak{S}$.‡

We shall develop only the theory of left ideals, as the theory of right ideals is obviously parallel.

In the usual manner§ we can show that every left ideal $\mathfrak{R}$ has a basis composed of $r \leqq n$ linearly independent numbers $\omega_1, \omega_2, \cdots, \omega_r$. That is, the numbers of $\mathfrak{R}$ are represented by the form

$$k_1\omega_1 + k_2\omega_2 + \cdots + k_r\omega_r$$

where $k_1, k_2, \cdots, k_r$ are rational integral variables. In particular the ω_i may be chosen in the form

$$\begin{aligned}
\omega_1 &= a_{11}e_1, \\
\omega_2 &= a_{21}e_1 + a_{22}e_2, \\
&\cdots\cdots\cdots\cdots \\
\omega_r &= a_{r1}e_1 + a_{r2}e_2 + \cdots + a_{rr}e_r
\end{aligned}$$

where each $a_{ii} > 0$.

---

* L. E. Dickson, *Algebras and their Arithmetics*, University of Chicago Press, 1923, p. 141. *Algebren und ihre Zahlentheorie*, Zürich, 1927, p. 155. These books will be referred to as Dickson I and Dickson II, respectively.

† Dickson I, p. 163, II, p. 212.

‡ This is more general than the definition of Speiser, which further specifies that the set $\mathfrak{R}$ be of order n, thus excluding what we shall call singular ideals. See Dickson II, p. 270.

§ Dickson II, p. 270. Landau, *Theorie der algebraischen Zahlen und der Ideale*, Leipzig, 1918, p. 29, Theorem 99.

This number r, which is evidently the number of linearly independent numbers in every basis of $\mathfrak{R}$, is called the *rank* of the ideal. The ideal is called *singular* or *non-singular* according as $r<n$ or $r=n$.

3. **Some properties of ideals.** Every number a satisfies its rank equation*

$$\omega^r + c_1\omega^{r-1} + \cdots + c_{r-1}\omega + N(a) = 0.$$

From the definition of ideal it follows that if a is in $\mathfrak{R}$, then

$$N(a) = (-a^{r-1} - c_1a^{r-2} - \cdots - c_{r-1})a$$

is also in $\mathfrak{R}$. Hence $\mathfrak{R}$ contains rational integers unless every number of $\mathfrak{R}$ is of norm 0.

Let p be the smallest positive integer in the ideal $\mathfrak{R}$. Let c be any rational integer in $\mathfrak{R}$. Then we may write $c=kp+r$ where $0\leqq r<p$. Since c and p are in $\mathfrak{R}$, so is $c-kp=r$. But p was minimal, so $r=0$ and $c=kp$. Hence *if not every number of $\mathfrak{R}$ is of norm 0, $\mathfrak{R}$ contains infinitely many rational integers, each an integral multiple of a smallest positive integer. If a is a number of $\mathfrak{R}$, so is $N(a)$.*

Let $\mathfrak{R}$ be an ideal containing a positive integer p. Since $\mathfrak{S}$ is of order n, it contains n linearly independent numbers $s_1, s_2, \cdots, s_n$. Then $\mathfrak{R}$ contains the linearly independent numbers $s_1p, s_2p, \cdots, s_np$. It follows that *if $\mathfrak{R}$ contains a positive integer, the numbers $\omega_1, \omega_2, \cdots, \omega_n$ of every basis are linearly independent and $\mathfrak{R}$ is of rank n.*

Let $\omega_1', \omega_2', \cdots, \omega_n'$ constitute a second basis for an ideal $\mathfrak{R}$ of rank n. There are rational integers a_{ij} and b_{ij} such that

$$\omega_i = \sum_k a_{ik}\omega_k', \quad \omega_i' = \sum_k b_{ik}\omega_k \qquad (i = 1,2,\cdots,n). \tag{3.1}$$

Then

$$\omega_i = \sum_{j,k} a_{ij}b_{jk}\omega_k,$$

and since the basal numbers are linearly independent,

$$\sum_j a_{ij}b_{jk} = \delta_{ik} \qquad (i,k = 1,2,\cdots,n)$$

so that (a_{rs}) is an integral matrix of determinant ± 1.

Conversely, if (a_{rs}) is an integral matrix of determinant ± 1, its inverse is likewise an integral matrix of determinant ± 1 and will serve as the matrix (b_{rs}) in (3.1). If the ω_i' constitute a basis for $\mathfrak{R}$, so do the ω_i as determined by

* Dickson I, p. 113.

(3.1). Thus *if* $\mathfrak{R}$ *is an ideal having a basis* $\omega_1, \omega_2, \cdots, \omega_n$ *composed of linearly independent numbers, every basis of* $\mathfrak{R}$ *is given by*

$$\omega_i' = \sum_j a_{ij}\omega_j \qquad (i = 1,2, \cdots, n)$$

where (a_{rs}) *is an integral matrix of determinant* ± 1.

If $\mathfrak{R}$ has a basis $\omega_1, \omega_2, \cdots, \omega_n$, we may write

$$\omega_i = \sum_j g_{ij}e_j \qquad (i = 1,2, \cdots, n), \tag{3.2}$$

where the g_{ij} are rational integers, in terms of the basal numbers $e_1, e_2, \cdots, e_n$ of $\mathfrak{S}$. We define the *norm* $N(\mathfrak{R})$ of the ideal $\mathfrak{R}$ to be the absolute value of the determinant $|g_{rs}|$. If $N(\mathfrak{R}) = 0$, the ω_i are linearly dependent and the rank r of $\mathfrak{R}$ is less than n; i.e., $\mathfrak{R}$ is singular.

Suppose that $\omega_1, \omega_2, \cdots, \omega_n$ constitute a linearly independent basis for $\mathfrak{R}$. If $\omega_1', \omega_2', \cdots, \omega_n'$ is any basis for $\mathfrak{R}$, we have

$$\omega_i' = \sum_j a_{ij}\omega_j = \sum_{j,k} a_{ij}g_{jk}e_k \equiv \sum_k g_{ik}'e_k \qquad (i = 1,2, \cdots, n)$$

where $|a_{rs}| = \pm 1$. Hence

$$(a_{rs})(g_{rs}) = (g_{rs}'),$$

which implies that

$$\text{absolute value } |g_{rs}'| = \text{absolute value } |g_{rs}| \neq 0,$$

so that the value of $N(\mathfrak{R})$ is independent of the basis chosen. In other words, $N(\mathfrak{R})$ is an invariant under change of basis and is therefore a positive integer intrinsically connected with the ideal $\mathfrak{R}$.

Let us now suppose that $\mathfrak{R}$ is an ideal for which $N(\mathfrak{R}) \neq 0$. From (3.2) we have

$$N(\mathfrak{R})e_i = \sum_j d_{ij}\omega_j \qquad (i = 1,2, \cdots, n)$$

where the d_{ij} are rational integers. Since e_1 is a principal unit, we see that $N(\mathfrak{R})$ is a positive integer in $\mathfrak{R}$. The norm of this positive integer is a non-zero integer in $\mathfrak{R}$, so that $\mathfrak{R}$ contains elements of norm not zero.

We may summarize the results of this paragraph in

THEOREM 3. *The following five statements are equivalent*:

(a) *Not every element of the ideal* $\mathfrak{R}$ *is of norm* 0.
(b) $\mathfrak{R}$ *contains rational integers.*
(c) *The numbers* $\omega_1, \omega_2, \cdots, \omega_n$ *of every basis of* $\mathfrak{R}$ *are linearly independent.*
(d) *The ideal* $\mathfrak{R}$ *is non-singular, i.e., of rank* n.
(e) *The norm* $N(\mathfrak{R})$ *of* $\mathfrak{R}$ *is not zero.*

We have shown that each statement implies the next following and that the last implies the first. Thus each of the five statements implies them all.

To see that ideals of norm zero actually exist, we need only to consider the principal ideal $(b]^*$ where $N(b)=0$, i.e., where b is a divisor of zero or zero itself. Then $\Delta(b)=0$, and for every number λ in $\mathfrak{S}$,

$$\Delta(\lambda b) = \Delta(\lambda)\cdot\Delta(b) = 0$$

so that $N(\lambda b)=0$ for every number λb in the ideal. Such ideals have none of the properties described in Theorem 3.

4. **A condition that n numbers constitute a basis for an ideal.** While it is true that every left ideal has a basis $\omega_1, \omega_2, \cdots, \omega_n$ composed of integral numbers, it does not follow that every n numbers of $\mathfrak{S}$ constitute a basis for a left ideal. We shall find necessary and sufficient conditions that n numbers $\omega_1, \omega_2, \cdots, \omega_n$ of $\mathfrak{S}$ shall form a basis for a left ideal.

Let us assume that $\omega_1, \omega_2, \cdots, \omega_n$ constitute a basis for a left ideal $\mathfrak{R}$, and that

$$\omega_i = \sum_j g_{ij}e_j \qquad (i = 1,2,\cdots,n). \tag{4.1}$$

Every number k of $\mathfrak{R}$ is of the form

$$k = \sum_i k_i\omega_i = \sum_{i,j} k_i g_{ij} e_j.$$

Every number s of $\mathfrak{S}$ is of the form

$$s = \sum_l s_l e_l.$$

Since sk is in $\mathfrak{R}$, there exist rational integers d_r such that

$$sk = \sum_{l,i,j} s_l k_i g_{ij} e_l e_j = \sum_{l,i,j,h} s_l k_i g_{ij} c_{ljh} e_h = \sum_{r,t} d_r g_{rt} e_t.$$

Since the basal numbers are linearly independent, this implies

$$\sum_{l,i,j} s_l k_i g_{ij} c_{ljt} = \sum_r d_r g_{rt} \qquad (t = 1,2,\cdots,n).$$

In particular there must exist rational integers d_r, which we shall call d_{pqr}, when $s_l=\delta_{lp}$ and $k_i=\delta_{iq}$. For these values we have

$$\sum_j g_{qj}c_{pjt} = \sum_r d_{pqr}g_{rt} \qquad (p,q,t = 1,2,\cdots,n). \tag{4.2}$$

---

* The notation is due to Speiser. See Dickson II, p. 271.

The existence of integers d_{pqr} satisfying (4.2) is in fact sufficient that the integral numbers $\omega_1, \omega_2, \cdots, \omega_n$ defined by (4.1) constitute a basis for a left ideal. For, let d_{pqr} and g_{qj} be any rational integers satisfying (4.2). Define the ω_i by (4.1). The set of numbers

$$k = \sum_i k_i \omega_i$$

where the k_i vary independently over $\mathfrak{S}$ is evidently closed under addition and subtraction. We can show that it is closed under multiplication on the left by any number s of $\mathfrak{S}$. In fact

$$sk = \sum_{l,i,j,r} s_l k_i g_{ij} c_{ljr} e_r = \sum_{l,i,r} s_l k_i \sum_j g_{ij} c_{ljr} e_r.$$

Hence by (4.2) we have

$$sk = \sum_{l,i,r} s_l k_i \left(\sum_s d_{lis} g_{sr} \right) e_r = \sum_{l,i,s} s_l k_i d_{lis} \omega_s,$$

which is obviously of the form $\sum_i k_i' \omega_i$.

Let us define the matrices

$$C_p = (c_{prs}) = \text{transpose } R_p, \; D_p = (d_{prs}), \; G = (g_{rs}) \qquad (p = 1, 2, \cdots, n).$$

We may now state

THEOREM 4.* *A necessary and sufficient condition in order that $\omega_1, \omega_2, \cdots, \omega_n$ constitute a basis for a left ideal is that there exist integral matrices $D_1, D_2, \cdots, D_n$ such that*

$$GC_p = D_p G \qquad (p = 1, 2, \cdots, n)$$

where G is the matrix (g_{rs}) of (4.1)

5. **An explicit basis for a principal ideal.** In particular every principal ideal $(d]$ has a basis $\omega_1, \omega_2, \cdots, \omega_n$. We have

$$d = \sum_i d_i e_i, \qquad \omega_i = \sum_j g_{ij} e_j \qquad (i = 1, 2, \cdots, n).$$

Let $s = \sum s_i e_i$ be any number of $\mathfrak{S}$. Then

$$sd = \sum_{i,j,k} s_i d_j c_{ijk} e_k$$

is in $(d]$, and hence must be representable in the form

$$sd = \sum_i r_i \omega_i = \sum_{i,j} r_i g_{ij} e_j.$$

* Poincaré, using the same correspondence between ideals and matrices, obtained a quite different condition that a matrix in canonical form correspond to an ideal. Bulletin de la Société Mathématique de France, vol. 13 (1885), p. 167.

It follows that for all rational integers s_i there must exist rational integers r_i such that

$$\sum_{i,j} s_i d_j c_{ijk} = \sum_i r_i g_{ik} \qquad (k = 1, 2, \cdots, n). \tag{5.1}$$

Conversely, since every linear combination of the ω_i is in $(d]$, it follows that for all integers r_i there must exist integers s_i such that (5.1) holds. These conditions together are necessary and sufficient in order that $\omega_1, \omega_2, \cdots, \omega_n$ form a basis for the principal left ideal $(d]$.

In particular, let us take $r_i = \delta_{hi}$, and denote the corresponding values of s_i by s_{hi}. Then (5.1) gives

$$\sum_{i,j} s_{hi} d_j c_{ijk} = \sum_i \delta_{hi} g_{ik} = g_{hk} \quad (h, k = 1, 2, \cdots, n). \tag{5.2}$$

Again, let us choose $s_i = \delta_{hi}$ in (5.1) and denote the corresponding values of r_i by r_{hi}. Then we have

$$\sum_{i,j} \delta_{hi} d_j c_{ijk} = \sum_j d_j c_{hjk} = \sum_i r_{hi} g_{ik} \qquad (h, k = 1, 2, \cdots, n). \tag{5.3}$$

From (5.2) we have, taking matrices,

$$G = (g_{rs}) = (s_{rs})\left(\sum d_j c_{rjs}\right) = (s_{rs}) S(d)$$

where $S(d)$ is the transposed second matrix* of d. In the same way we obtain from (5.3)

$$S(d) = (r_{rs}) G.$$

Every number of the ideal $(d]$ is of the form

$$\sum_i a_i \omega_i = \sum_{i,j} a_i g_{ij} e_j = \sum_{i,k,j} a_i s_{ik} \sigma_{kj} e_j,$$

where $S(d) = (\sigma_{rs})$, and conversely every number of the form $\sum b_i \sigma_{ij} e_j$ can be written

$$\sum_{i,j} b_i \sigma_{ij} e_j = \sum_{i,k,j} b_i \sigma_{ik} g_{kj} e_j = \sum_{i,k} b_i \sigma_{ik} \omega_k$$

and is therefore in $(d]$. Hence the numbers $\sum \sigma_{ij} e_j$ constitute a basis for $(d]$.

It follows from the definition of norm that $N((d]) = $ *absolute value* $S(d) = $ *absolute value* $\Delta'(d)$. We have now proved

THEOREM 5.1. *The principal left ideal* $(d]$ *has a basis* $\omega_1, \omega_2, \cdots, \omega_n$ *where*

$$\omega_i = \sum_j \sigma_{ij} e_j, \quad (\sigma_{rs}) = S(d);$$

the norm of $(d]$ *is the absolute value of* $\Delta'(d)$.

* Dickson I, p. 86, II, p. 35. It is easy to show that d is an integral number if and only if $S(d)$ has integral elements.

The conditions for associativity may be written

$$\sum_j c_{qkj}c_{pjs} = \sum_j c_{pqj}c_{jks}.$$

Let us multiply by d_k and sum for k:

$$\sum_{j,k} d_k c_{qkj} c_{pjs} = \sum_{j,k} c_{pqj} d_k c_{jks},$$

which may be written

$$S(d)C_p = C_p S(d) \qquad (p = 1, 2, \cdots, n)$$

where $d = \sum d_i e_i$. This is exactly (4.2) with g_{ij} replaced by σ_{ij} and d_{pqr} replaced by c_{pqr}.

THEOREM 5.2. *For every number d in $\mathfrak{S}$, $S(d)$ is commutative with every matrix C_p; and if $G = S(d)$ in Theorem 4, then $D_p = C_p$ for every p.*

We shall prove later (Theorem 11) that every integral matrix commutative with every C_p is the transposed second matrix $S(d)$ of some number d of $\mathfrak{S}$.

6. **Equivalent ideals.** Let $\mathfrak{R}$ be a non-singular left ideal with basis $\omega_1, \omega_2, \cdots, \omega_n$, and let s be a number of $\mathfrak{S}$. If we set

$$\omega_i = \sum_j g_{ij} e_j, \qquad s = \sum_q s_q e_q,$$

we have

$$\omega_i s \equiv \sum_j g'_{ij} e_j = \sum_p g_{ip} e_p \sum_q s_q e_q$$
$$= \sum_{p,q,r} g_{ip} s_q c_{pqr} e_r.$$

Hence

$$g'_{ij} = \sum_{p,q} g_{ip} s_q c_{pqj} = \sum_p g_{ip} \sum_q s_q c_{pqj} = \sum_p g_{ip} \sigma_{pj}.$$

Taking matrices, we have

$$G' = GS(s).$$

But

$$GC_p = D_p G, \qquad S(s)C_p = C_p S(s),$$

so that

$$G'C_p = GS(s)C_p = GC_p S(s) = D_p GS(s) = D_p G'.$$

Therefore $\omega_1 s, \omega_2 s, \cdots, \omega_n s$ form a basis for an ideal which we may call $\mathfrak{R}'$.

Furthermore, we obtain the same ideal $\mathfrak{R}'$ irrespective of the basis of $\mathfrak{R}$ with which we start. Using any other basis for $\mathfrak{R}$, we should have obtained, instead of G',

$$G'' = AGS(s)$$

where A is an integral matrix of determinant ± 1 (§3). But $G''=AG'$ corresponds to the same ideal $\mathfrak{R}'$ as does G'. For a non-singular ideal $\mathfrak{R}$ we may call the uniquely existing ideal having the basis $(\omega_1 s, \omega_2 s, \cdots, \omega_n s)$ where $(\omega_1, \omega_2, \cdots, \omega_n)$ is any basis for $\mathfrak{R}$, the ideal $\mathfrak{R}s$.

Two non-singular ideals $\mathfrak{R}_1$ and $\mathfrak{R}_2$ are, according to analogy with the usual definition in algebraic number theory, called *equivalent* if there exist integral numbers s_1 and s_2 of $\mathfrak{S}$, neither of norm 0, such that

$$\mathfrak{R}_1 s_1 = \mathfrak{R}_2 s_2.$$

Let G_1 and G_2 be the matrices corresponding to particular bases of $\mathfrak{R}_1$ and $\mathfrak{R}_2$ respectively. The condition for equivalence becomes

$$G_1 S(s_1) = A G_2 S(s_2)$$

where A is an integral matrix of determinant ± 1.

The following theorem is important in showing that the concept of ideal class applies to all semi-simple domains and in pointing to a more comprehensive definition which is applicable to singular ideals as well.

THEOREM 6. *A necessary and sufficient condition that two non-singular ideal matrices G_1 and G_2 be equivalent is that the corresponding sets of matrices D_{1p}, D_{2p} satisfying the equations*

$$G_1 C_p = D_{1p} G_1, \quad G_2 C_p = D_{2p} G_2 \qquad (p = 1, 2, \cdots, n)$$

respectively, be similar—i.e., that $D_{1p} = A D_{2p} A^{-1}$ for $p = 1, 2, \cdots, n$, where A is an integral matrix of determinant ± 1.

First, let us suppose that G_1 and G_2 are equivalent. We have

$$G_1 C_p = D_{1p} G_1, \qquad G_2 C_p = D_{2p} G_2,$$
$$G_1 S(s_1) = A G_2 S(s_2), \qquad |A| = \pm 1$$

where $S(s_1)$ and $S(s_2)$ are each, by Theorem 5.2, commutative with every C_p. Then

$$G_1 S(s_1) C_p = G_1 C_p S(s_1) = D_{1p} G_1 S(s_1),$$
$$A G_2 S(s_2) C_p = A G_2 C_p S(s_2) = A D_{2p} G_2 S(s_2)$$
$$= A D_{2p} A^{-1} A G_2 S(s_2).$$

Therefore

$$D_{1p} G_1 S(s_1) = A D_{2p} A^{-1} G_1 S(s_1).$$

Since both G_1 and $S(s_1)$ are non-singular, we have

$$D_{1p} = A D_{2p} A^{-1} \qquad (p = 1, 2, \cdots, n).$$

Secondly, suppose that the sets of matrices D_{1p} and D_{2p} are similar:

$$G_1C_p = D_{1p}G_1, \quad G_2C_p = D_{2p}G_2, \quad D_{1p} = AD_{2p}A^{-1}, \quad |A| = \pm 1.$$

Then

$$G_1C_pG_1^{-1} = D_{1p} = AD_{2p}A^{-1} = AG_2C_pG_2^{-1}A^{-1}.$$

We multiply on the left by $G_2^{-1}A^{-1}$ and on the right by G_1, obtaining

$$G_2^{-1}A^{-1}G_1C_p = C_pG_2^{-1}A^{-1}G_1.$$

Let T be the scalar matrix each of whose diagonal elements is $|G_2|$. Then $TG_2^{-1}A^{-1}G_1$ is an integral matrix and, since it is commutative with each C_p, it is the second matrix $S(s)$ of some element s of $\mathfrak{S}$.* Then

$$TG_2^{-1}A^{-1}G_1 = S(s),$$

and since T is commutative with every matrix,

$$G_1T = AG_2S(s).$$

Moreover, $T = S(|G_2|)$, so that the ideals $\mathfrak{R}_1$ and $\mathfrak{R}_2$ are in fact equivalent.

7. **Ideal matrices.** An essential point in our proof of Theorem 6 was that $|G_1| \neq 0$. We were therefore unable to consider equivalence of singular ideals. Moreover, the transitive character of equivalence was not apparent from its definition. We now proceed along a line suggested by this theorem but somewhat broader.

We assume a semi-simple rational algebra $\mathfrak{A}$, and a set of integral numbers $\mathfrak{S}$ of order n, the basal numbers being chosen so that the constants c_{ijk} of multiplication are rational integers. As before, we define the matrices $C_p = (c_{prs})$ for $p = 1, 2, \cdots, n$, where c_{prs} is the element in row r and column s. Let $D_p = (d_{prs})$ be any set of n integral matrices. All the integral matrices G which satisfy the equations

$$GC_p = D_pG \qquad (p = 1, 2, \cdots, n) \tag{7.1}$$

will be said to constitute a *minor class* of *ideal matrices*, and the set of matrices $D_1, D_2, \cdots, D_n$ will be called a set of corresponding *class matrices*. The zero matrix at least will satisfy (7.1) no matter how the class matrices may be chosen.

All the matrices of a minor class constitute a modul. In fact, if $G_1, G_2, \cdots, G_p$ are ideal matrices of the same minor class and if $k_1, k_2, \cdots, k_p$ are rational integers, then

$$G = k_1G_1 + k_2G_2 + \cdots + k_pG_p$$

* Theorem 11, to follow.

is an ideal matrix of the same minor class. We shall show that if $\mathfrak{A}$ is semi-simple there exist $r \leqq n$ linearly independent matrices $B_1, B_2, \cdots, B_r$ which constitute a basis for the ideal matrices of the minor class.

LEMMA 7. *If $\mathfrak{A}$ is a semi-simple algebra, new basal numbers for $\mathfrak{S}$ and new constants of multiplication c'_{ijk} can be so chosen that $|c'_{rs1}| \neq 0$.*

Let us suppose that our basal numbers are chosen as in §2, and that the c_{ijk} are defined by

$$e_i e_j = \sum_k c_{ijk} e_k \qquad (i,j = 1,2, \cdots, n). \tag{7.2}$$

Since e_1 is a principal unit, $c_{1jk} = c_{j1k} = \delta_{jk}$. If we apply to the basal numbers the transformation

$$e_i = \sum_j a_{ij} e'_j, \qquad a = |a_{rs}| \neq 0, \tag{7.3}$$

we get from (7.2):

$$\sum_{p,q} a_{ip} a_{jq} e'_p e'_q = \sum_{r,s} c_{ijr} a_{rs} e'_s = \sum_{p,q,t} a_{ip} a_{jq} c'_{pqt} e'_t$$

so that

$$\sum_{p,q} a_{rp} a_{sq} c'_{pqj} = \sum_i c_{rsi} a_{ij} \qquad (r,s,j = 1,2, \cdots, n), \tag{7.4}$$

where the c'_{pqs} are defined as in (7.2) with each letter primed. For a fixed j, let us form the determinant whose element in row r and column s is (7.4):

$$a^2 |c'_{rsj}| = \left| \sum_i a_{ij} c_{rsi} \right| \qquad (j = 1,2, \cdots, n). \tag{7.5}$$

In (7.3) let us now choose

$$a_{ij} = \tau_{ij} \equiv \sum_{r,k} c_{ijr} c_{rkk}.$$

If $\mathfrak{A}$ is semi-simple, $a = |\tau_{rs}| \equiv d \neq 0$,* and

$$a_{r1} = \sum_{h,k} c_{r1h} c_{hkk} = \sum_{h,k} \delta_{hr} c_{hkk} \equiv \sum_k c_{rkk}$$

so that

$$\sum_r a_{r1} c_{ijr} = \sum_{r,k} c_{ijr} c_{rkk} = \tau_{ij}.$$

---

* Dickson I, p. 108.

From (7.5) with $j=1$ we obtain

$$d^2 \left| c'_{rs1} \right| = d$$

so that $|c'_{rs1}| \neq 0$.

In making this transformation we may have lost the property that the basal numbers of $\mathfrak{A}$ form a basis for the set $\mathfrak{S}$ of integral numbers. We must now restore this property.

Bring* the fractions c'_{ijk} to a common denominator δ, write $\delta c'_{ijk} = h_{ijk}$ where the h_{ijk} are all integers. Set $\epsilon_i = \delta e'_i$. Then

$$\epsilon_i \epsilon_j = \delta^2 e'_i e'_j = \delta^2 \sum_k c'_{ijk} e'_k = \sum_k h_{ijk} \epsilon_k.$$

We use the new basal numbers ϵ_i whose constants h_{ijk} of multiplication are rational integers. We note that

$$|h_{rs1}| = |\delta c'_{rs1}| = \delta^n |c'_{rs1}| \neq 0.$$

Proceeding according to the method of Dickson,† we see that every element x of $\mathfrak{S}$ can be put into the form

$$x = \frac{x_1}{D}\epsilon_1 + \frac{x_2}{D}\epsilon_2 + \cdots + \frac{x_n}{D}\epsilon_n \tag{7.6}$$

where

$$D = \left| \sum_{i,k} h_{rsi} h_{ikk} \right| \neq 0,$$

and where the x_i are rational integers. Of all numbers x in $\mathfrak{S}$ having $x_1 = x_2 = \cdots = x_{r-1} = 0$ but $x_r \neq 0$, choose one having $x_r > 0$ and minimal for ϵ'_r. If there is no x of this type having $x_r \neq 0$, choose $\epsilon'_r = 0$. We have

$$\begin{aligned} D\epsilon'_1 &= b_{11}\epsilon_1 + b_{12}\epsilon_2 + \cdots + b_{1n}\epsilon_n, \\ D\epsilon'_2 &= \quad\ b_{22}\epsilon_2 + \cdots + b_{2n}\epsilon_n, \\ &\cdots\cdots\cdots\cdots \\ D\epsilon'_n &= \qquad\qquad\quad b_{nn}\epsilon_n. \end{aligned} \tag{7.7}$$

Now $(\epsilon'_1, \epsilon'_2, \cdots, \epsilon'_n)$ form a basis for $\mathfrak{S}$. For, let

$$x = \frac{1}{D}(x_1\epsilon_1 + x_2\epsilon_2 + \cdots + x_n\epsilon_n)$$

---

* Dickson I, pp. 161–162.

† Dickson I, p. 162.

be any number of $\mathfrak{S}$. Set

$$x_1 = q_1 b_{11} + r_1 \qquad (0 \leqq r_1 < b_{11}).$$

Then

$$x - q_1\epsilon_1' = \frac{1}{D}(r_1\epsilon_1 + x_2'\epsilon_2 + \cdots + x_n'\epsilon_n)$$

has its coefficient of ϵ_1 less than b_{11} which was minimal, so $r_1 = 0$. Similarly, set

$$x_2' = q_2 b_{22} + r_2 \qquad (0 \leqq r_2 < b_{22}).$$

Then $x - q_1\epsilon_1' - q_2\epsilon_2'$ lacks the basal numbers ϵ_1 and ϵ_2. Proceeding in this way, we have after n steps:

$$x = q_1\epsilon_1' + q_2\epsilon_2' + \cdots + q_n\epsilon_n'$$

where the q_i are rational integers. Since $\mathfrak{S}$ is of order n, we now know that $b_{ii} \neq 0$ for every i.

Since the numbers $\epsilon_1', \epsilon_2', \cdots, \epsilon_n'$ form a basis for $\mathfrak{S}$, we know that the new constants h'_{jk} of multiplication are rational integers. Transformation (7.7) can be written

$$\epsilon_i = \frac{D}{b}\sum_j B_{ji}\epsilon_j' \qquad (i = 1, 2, \cdots, n)$$

where $b = |b_{rs}|$ and B_{rs}, the cofactor of b_{rs}, is zero for $r < s$. From (7.5) we have

$$a^2 \left| h'_{rs1} \right| = \left| \sum_i a_{i1} h_{rsi} \right|$$

with $a_{ij} = (1/b)DB_{ji} = 0$ for $i > j$. Hence

$$a^2 \left| h'_{rs1} \right| = a_{11}^n \left| h_{rs1} \right| \neq 0,$$

and therefore $|h'_{rs1}| \neq 0$.

Since the basal numbers $e_1, e_2, \cdots, e_n$ with which we started and the numbers $\epsilon_1', \epsilon_2', \cdots, \epsilon_n'$ each form a basis for the set $\mathfrak{S}$, we know that there are transformations with rational integral coefficients and of determinants ± 1 carrying each set of basal numbers into the other set.

8. **Effect of change of basis on the fundamental matrices.** Let us make a change of basis (7.3) where the a_{rs} are rational integers of determinant ± 1. By (7.4) we have

$$\sum_r c_{pir} a_{rt} = \sum_{s,q} a_{ps} a_{iq} c'_{sqt}.$$

Multiply by the cofactor A_{jt} of a_{jt} and sum for t. Then

$$ac_{pij} = \sum_{q,t} a_{iq}\left[\sum_{s} a_{ps}c'_{sqt}\right]A_{jt}.$$

Passing to matrices, we have

$$C_p = A\left[\sum_{s} a_{ps}C'_s\right]A^{-1} \qquad (p = 1,2,\cdots,n)$$

where $A \equiv (a_{rs})$ as our induced transformation on the matrices C_p.

We shall more generally define the matrices D'_p, which will be said to correspond to the matrices D_p under transformation (7.3) of determinant ± 1, as the solutions of the equations

$$D_p = A\left[\sum_{s} a_{ps}D'_s\right]A^{-1} \qquad (p = 1,2,\cdots,n). \tag{8.1}$$

The matrices D'_p are evidently integral if the matrices D_p are integral. Our definition is justified by the following result:

LEMMA 8.1. *If G is an ideal matrix satisfying the equations*

$$GC_p = D_pG \qquad (p = 1,2,\cdots,n),$$

then $G' = A^{-1}GA$ satisfies the equations

$$G'C'_p = D'_pG' \qquad (p = 1,2,\cdots,n)$$

where C'_p and D'_p are given by (8.1).

From (8.1) and the equations $GC_p = D_pG$, we have

$$GA\left[\sum_{s} a_{ps}C'_s\right]A^{-1} = A\left[\sum_{s} a_{ps}D'_s\right]A^{-1}G,$$

and therefore

$$A^{-1}GA\left[\sum_{s} a_{ps}C'_s\right] = \left[\sum_{s} a_{ps}D'_s\right]A^{-1}GA.$$

If we multiply by the cofactor A_{pr} of a_{pr} and sum for p, we obtain

$$A^{-1}GAC'_r = D'_rA^{-1}GA,$$

which proves the lemma.

LEMMA 8.2. *If the basis of $\mathfrak{S}$ is so chosen that $|c_{r\,s1}| \neq 0$, then an ideal matrix G whose first column consists exclusively of zeros is a zero matrix.*

From our definition of ideal matrix we have

$$\sum_k g_{rk}c_{pks} = \sum_k d_{prk}g_{ks} \qquad (p,r,s = 1,2,\cdots,n).$$

If $g_{k1}=0$ for every k, then

$$\sum_k g_{rk}c_{pk1} = 0 \qquad (r,p = 1,2,\cdots,n).$$

Taking matrices, we have

$$G(c_{sr1}) = 0.$$

Since the second matrix is non-singular, G must be of rank 0.

9. **A fundamental theorem.** Let $\mathfrak{A}$ be a semi-simple algebra, and let $e_1, e_2, \cdots, e_n$ be a basis for a set $\mathfrak{S}$ of integral numbers of $\mathfrak{A}$, where e_1 is a principal unit. We transform by a matrix A of determinant ± 1 to another basis $e_1', e_2', \cdots, e_n'$ for which $|c'_{rs1}| \neq 0$. Every ideal matrix G is thereby transformed into an ideal matrix $G'=A^{-1}GA$, and inversely $G=AG'A^{-1}$, so that there is a one-to-one correspondence between the ideal matrices G and G'.

If G' is an ideal matrix of a certain minor class, so is $-G'$. If in the minor class defined by the class matrices $D_1', D_2', \cdots, D_n'$ (§7), there are ideal matrices $G'=(g_{rs})$ in which $g_{11}\neq 0$, define as B_1' one such matrix in which $g_{11}>0$ and is minimal. If no such matrix exists, set $B_1'=0$, which surely is in the minor class. If there are in the minor class matrices for which $g_{11}=g_{21}=\cdots=g_{k-1,1}=0$ but $g_{k1}\neq 0$, define as $B_k'=(b_{krs})$ one such matrix in which $g_{k1}=b_{kk1}>0$ and is minimal, otherwise set $B_k'=0$. Then B_1', $B_2', \cdots, B_n'$ form a basis for the matrices of the minor class.

For, let $G'=(g_{rs})$ be any matrix of the minor class. Set

$$g_{11} = h_1 b_{111} + r_1 \qquad (0 \leqq r_1 < b_{111}).$$

Then $G'-h_1B_1'=(g_{rs}')$ is an ideal matrix of the minor class having as its first element r_1. But b_{111} was minimal, so $r_1=0$. Now set

$$g_{21}' = h_2 b_{221} + r_2 \qquad (0 \leqq r_2 < b_{221}).$$

Similarly

$$G' - h_1B_1' - h_2B_2'$$

has two zeros in its first column. Proceeding in this way, we find after n steps that

$$G' - h_1B_1' - h_2B_2' - \cdots - h_nB_n'$$

is an ideal matrix of the minor class whose first column consists exclusively of zeros, and which by Lemma 8.2 is therefore the zero matrix. Thus

$$G' = h_1B_1' + h_2B_2' + \cdots + h_nB_n'$$

where $h_1, h_2, \cdots, h_n$ are rational integers.

Let us now transform back to our original basis. We have $G=AG'A^{-1}$, and we define $B_i \equiv AB_i'A^{-1}$, so that every ideal matrix G is expressible in the form

$$G = h_1B_1 + h_2B_2 + \cdots + h_nB_n$$

where the $h_1, h_2, \cdots, h_n$ are rational integers. Conversely, every such matrix is in the minor class. We now have

THEOREM 9. *Relative to every basis $e_1, e_2, \cdots, e_n$ for a set $\mathfrak{S}$ of integral elements of a semi-simple rational algebra $\mathfrak{A}$, every minor class of ideal matrices has a basis composed of n matrices $B_1, B_2, \cdots, B_n$ such that the totality of ideal matrices of the minor class is given by*

$$h_1B_1 + h_2B_2 + \cdots + h_nB_n$$

where $h_1, h_2, \cdots, h_n$ are independent rational integral variables.

10. **Rank of a minor class.** Suppose that as in the preceding paragraph we have a set $\mathfrak{S}$ of integral numbers with basis so chosen that $|c_{rs1}'| \neq 0$. Then we have seen that every minor class has a basis $B_1', B_2', \cdots, B_n'$ such that in each matrix $B_k' \equiv (b_{krs})$ we have $b_{kr1}=0$ for $r<k$, and either $b_{kk1}>0$ and minimal, or $B_k'=0$. Suppose that the B_k' are linearly dependent:

$$d_1B_1' + d_2B_2' + \cdots + d_nB_n' = 0,$$

where we may assume that the d_i are rational integers not all zero. Considering only elements in the first columns, we have

$$d_1b_{1r1} + d_2b_{2r1} + \cdots + d_nb_{nr1} = 0 \quad (r = 1,2, \cdots, n).$$

Suppose that $d_1=d_2=\cdots=d_{p-1}=0$ while $d_p \neq 0$. Then we have $d_pb_{pp1}=0$ so that $b_{pp1}=0$ and hence $B_p'=0$.

If there is a dependence relation among the remaining $n-1$ basal matrices B_i', we may repeat the argument and show that another one is zero. We finally reach a point where all the basal matrices which are not zero are linearly independent. The number r of linearly independent matrices in a basis is called the *rank* of the minor class, and if $r<n$ the class is called *singular*. When we transform to another basis for $\mathfrak{S}$, we see from the relation $B_i=AB_i'A^{-1}$ that the rank is preserved.

Just as in the case of change of basis of an ideal, it can be shown that, relative to the same basis for $\mathfrak{S}$, the most general transformation from one linearly independent basis to another is given by

$$B_i' = \sum_{j=1}^{r} a_{ij}B_j \qquad (i = 1,2, \cdots, r)$$

where (a_{rs}) is an integral matrix of determinant ± 1. The rank is preserved under such transformations also.

11. **The principal minor class.** All matrices commutative with the fundamental matrices C_p evidently constitute a minor class, called the *principal* minor class. The members of this class we shall call *principal* ideal matrices. We shall now prove the result required to complete the proof of Theorem 6:

THEOREM 11. *Every principal ideal matrix is the transposed second matrix $S(k)$ of some integral number k of $\mathfrak{S}$, and conversely.*

The conditions for associativity may be written*

$$\sum_h c_{rih}c_{phs} = \sum_h c_{prh}c_{his}$$

or

$$(c_{ris})C_p = C_p(c_{ris}).$$

That is, the matrices $(c_{ris}) = S_i = S(e_i)$ belong to the principal minor class. Let $k = k_1e_1 + k_2e_2 + \cdots + k_ne_n$ be any integral number. Then

$$S(k) = k_1S_1 + k_2S_2 + \cdots + k_nS_n$$

where the k_i are rational integers, and therefore $S(k)$ is in the principal minor class. This proves the converse.

We consider now the set of all matrices G which are commutative with every C_p. We have seen that every such minor class has a basis $B_1, B_2, \cdots, B_n$. Since the matrices S_i are in this set, we have

$$S_i = \sum_j a_{ij}B_j \qquad (i = 1,2, \cdots, n)$$

where the a_{ij} are rational integers. Now $|a_{rs}| \neq 0$, since the S_i are linearly independent. Therefore we can solve these equations for the B_i, obtaining

$$B_i = \sum_j r_{ij}S_j = S(r_i) \qquad (i = 1,2, \cdots, n)$$

where each r_{ij} is rational, and $r_i = r_{i1}e_1 + r_{i2}e_2 + \cdots + r_{in}e_n$ is a number of the algebra $\mathfrak{A}$. But $S(r_i) = B_i$ is an integral matrix, and hence by the footnote to §5 each r_i is an integral number and therefore the r_{ij} are rational integers. Then every matrix

$$G = h_1B_1 + h_2B_2 + \cdots + h_nB_n$$

can be written

$$G = S(k), \qquad k = \sum_{i,j} h_ir_{ij}e_j,$$

* Dickson I, p. 92.

where the h_i and r_{ij} are rational integers, so that k is a number of $\mathfrak{S}$. This completes the proof of the theorem.

12. **The class number.** We shall now establish a few properties of minor classes.

THEOREM 12.1. *If G is an ideal matrix of the minor class $\mathfrak{k}$, and if P is a principal ideal matrix, then GP is an ideal matrix of the minor class $\mathfrak{k}$.*

Let $D_1, D_2, \cdots, D_n$ be the class matrices defining the minor class $\mathfrak{k}$. Then

$$GC_p = D_pG, \qquad PC_p = C_pP \qquad (p = 1,2, \cdots, n).$$

Therefore

$$GPC_p = GC_pP = D_pGP,$$

which proves the theorem.

THEOREM 12.2. *If the minor class $\mathfrak{k}$ contains one non-singular ideal matrix G, then the n basal matrices of $\mathfrak{k}$ are linearly independent, and $\mathfrak{k}$ is of rank n.*

Let G be a non-singular matrix of $\mathfrak{k}$. Let $P_1, P_2, \cdots, P_n$ be linearly independent matrices of the principal class. Then $GP_1, GP_2, \cdots, GP_n$ are linearly independent matrices of class $\mathfrak{k}$, for if there were a dependence relation

$$\sum_i d_iGP_i = G\sum_i d_iP_i = 0, \tag{12.1}$$

where G is a non-singular matrix, we should have $\sum d_iP_i=0$, contrary to assumption. Let $B_1, B_2, \cdots, B_n$ be a basis for $\mathfrak{k}$. Then

$$GP_i = \sum_j b_{ij}B_j \qquad (i = 1,2, \cdots, n),$$

and since the GP_i are linearly independent, so are the basal matrices B_j. Thus the rank of $\mathfrak{k}$ is n.

THEOREM 12.3. *If $\mathfrak{A}$ is a division algebra, every minor class except the zero class is non-singular.*

Since $\sum d_iP_i$ is the transposed second matrix of a number of $\mathfrak{A}$, it is either of rank n or of rank 0. Thus in (12.1) either $G=0$, or else $d_1=d_2=\cdots=d_n=0$ and the matrices GP_i are linearly independent.

THEOREM 12.4. *If two minor classes contain the same non-singular ideal matrix, the classes coincide.*

Suppose that

$$GC_p = D_{1p}G, \quad GC_p = D_{2p}G, \quad |G| \neq 0 \qquad (p = 1,2, \cdots, n).$$

Then $D_{1p}G = D_{2p}G$, and since G is non-singular, $D_{1p} = D_{2p}$ for every p. Thus the minor classes coincide.

The theorem is not true with the omission of the word "non-singular." Thus the zero matrix is common to every minor class.

In general we shall not expect the number of minor classes to be finite. Thus $D_1, D_2, \cdots, D_n$ may be taken as perfectly arbitrary integral matrices, and the equations $GC_p = D_pG$ will be satisfied by the zero matrix at least. Moreover, if there be matrices satisfying a relation $GC_p = D_pG$, then

$$AGC_p = AD_pA^{-1}AG,$$

so that the matrices AD_pA^{-1} define a minor class in general distinct from the given class.

Two minor classes whose class matrices are connected by a relation

$$D_p' = AD_pA^{-1}, \quad |A| = \pm 1 \qquad (p = 1,2, \cdots, n)$$

will be called *similar* minor classes. We now define the (left) *class number h* of the set $\mathfrak{S}$ of integral numbers of $\mathfrak{A}$ as the (cardinal) number of dissimilar non-singular minor classes of (left) ideal matrices.

It is evident from Theorem 6 that when $\mathfrak{A}$ is an algebraic field, h becomes the ordinary class number of the field. All ideal matrices corresponding to principal ideals belong to minor classes which are similar to the principal minor class. We have therefore, without using the concept of ideal multiplication, succeeded in generalizing to sets $\mathfrak{S}$ of semi-simple algebras the concept of ideal class in a satisfactory manner. For instance, we may prove in the usual manner

THEOREM 12.5. *A necessary and sufficient condition in order that every pair of numbers of $\mathfrak{S}$ may possess a greatest common right divisor expressible linearly in terms of the numbers is that the left class number h of $\mathfrak{S}$ be* 1.

13. **The density of ideal matrices.** It is recognized that the addition of Dedekind ideals cannot be defined in any useful way, because of the fact that associated numbers correspond to the same principal ideal. This is not true of ideal matrices, however, and we have a satisfactory additive theory within each minor class.

Let $B_1, B_2, \cdots, B_n$ be a set of basal matrices for a non-singular minor class $\mathfrak{k}$. Every ideal matrix of $\mathfrak{k}$ has the form

$$G = a_1B_1 + a_2B_2 + \cdots + a_nB_n$$

where the a's are rational integers, and conversely every such G is in $\mathfrak{k}$. Since the B_i are linearly independent, this representation is unique. Addition and scalar multiplication within the minor class $\mathfrak{k}$ are defined as in the theory of linear algebras, and follow the usual laws.

We have seen that $S_1, S_2, \cdots, S_n$ constitute a basis for the principal minor class (§11). The numbers

$$a = a_1e_1 + a_2e_2 + \cdots + a_ne_n$$

are in one-to-one correspondence with the principal ideal matrices

$$S(a) = a_1S_1 + a_2S_2 + \cdots + a_nS_n,$$

which in turn are in one-to-one correspondence with the ideal matrices G of each class $\mathfrak{k}$, and this correspondence is preserved under addition and scalar multiplication. Thus we have

THEOREM 13. *The ideal matrices of every non-singular minor class $\mathfrak{k}$ are in one-to-one correspondence with the numbers of the set $\mathfrak{S}$. This correspondence is preserved under addition and scalar multiplication.*

OHIO STATE UNIVERSITY,
COLUMBUS, OHIO

ON THE PADÉ APPROXIMANTS ASSOCIATED WITH THE CONTINUED FRACTION AND SERIES OF STIELTJES*

BY

HUBERT S. WALL

INTRODUCTION

The well known series of Stieltjes is a power series with real coefficients:

$$\frac{c_0}{z} - \frac{c_1}{z^2} + \frac{c_2}{z^3} - \cdots \tag{1}$$

such that the persymmetric determinants

$$A_n \equiv \begin{vmatrix} c_0 & c_1, \cdots, c_{n-1} \\ c_1 & c_2, \cdots, c_n \\ \cdot & \cdot \quad \cdot \\ c_{n-1}, & c_n, \cdots, c_{2n-2} \end{vmatrix}, \qquad B_n \equiv \begin{vmatrix} c_1, & c_2 & , \cdots, c_n \\ c_2, & c_3 & , \cdots, c_{n+1} \\ \cdot & \cdot & \cdot \\ c_n, & c_{n+1}, & \cdots, c_{2n-1} \end{vmatrix}$$

($n=1, 2, 3, \cdots, A_0 \equiv 1, B_0 \equiv 1$) are all positive. Stieltjes† connected the series (1) with a continued fraction

$$\frac{1}{a_1 z} + \frac{1}{a_2} + \frac{1}{a_3 z} + \cdots \tag{2}$$

in which all the a_i are real and positive, and are uniquely determined by the c_i in accordance with the relations

$$a_{2n} = A_n^2/(B_n B_{n-1}), \quad a_{2n+1} = B_n^2/(A_n A_{n+1}). \tag{3}$$

Conversely, the positive real numbers a_i ($i=1, 2, 3, \cdots$) uniquely determine coefficients c_i of a Stieltjes power series. Thus to every Stieltjes continued fraction (2) with positive a_i there corresponds a Stieltjes series (1) and vice versa.

In the work of Stieltjes the convergents of (2) separate into two sets, the odd and the even convergents. When the series‡ $\sum a_i$ is divergent, these

* Presented to the Society, April 16, 1927; received by the editors June 11, 1928. This paper is essentially a thesis prepared at the suggestion of Professor E. B. Van Vleck at the University of Wisconsin.

† *Recherches sur les fractions continues*, Annales de Toulouse, vol. 8, J, pp. 1–122, and vol. 9, A, pp. 1–47, 1894–95. Published also in vol. 32 of the *Mémoires présentés à l'Académie des Sciences de l'Institut National de France.*

‡ Here and henceforth we write $\sum$ in place of $\sum_{i=1}^{\infty}$.

two sets of alternate convergents have one and the same limit which is a function analytic throughout the complex plane, excepting the whole or a part of the negative real axis. When, on the other hand, the series $\sum a_i$ converges, these two sets have distinct limits p_1/q_1 and p/q respectively, which are meromorphic functions having only simple poles all of which are real and non-positive. The numerators and denominators p, p_1, q, q_1 are analytic functions between which there is the relation

$$p_1 q - p q_1 = +1. \tag{4}$$

In the case of a convergent Stieltjes series (1) the former alternative is always realized, but for a divergent series either alternative may be realized. Thus for some divergent series the continued fraction of Stieltjes picks out two functional equivalents, while for others only a single equivalent is yielded.

In this article it will be convenient to replace $1/z$ by z in (1) and (2) and then drop the unessential factor z. We then have the Stieltjes series

$$c_0 - c_1 z + c_2 z^2 - c_3 z^3 + \cdots \tag{5}$$

with the same inequalities:

$$A_n > 0, \quad B_n > 0, \tag{6}$$

while the continued fraction becomes

$$\cfrac{1}{a_1 + \cfrac{z}{a_2 + \cfrac{z}{a_3 + \cfrac{z}{a_4 + \cdots}}}}. \tag{7}$$

When $\sum a_i$ is convergent, the two sets of alternate convergents have limits which are distinct and meromorphic over the entire plane *except at the origin*, while if $\sum a_i$ is divergent both sets have the same limit which is a function analytic over the entire plane except the whole or a part of the negative real axis.

With any power series

$$k_0 + k_1 z + k_2 z^2 + \cdots \qquad (k_0 \neq 0), \tag{8}$$

Padé connects a table of approximants*

	$n = 0$	$n = 1$	
$m = 0$	$\frac{N_{0,0}}{D_{0,0}} = k_0,$	$\frac{N_{0,1}}{D_{0,1}} = k_0 + k_1 z,$	$\cdots$
$m = 1$	$\frac{N_{1,0}}{D_{1,0}},$	$\frac{N_{1,1}}{D_{1,1}},$	$\cdots$
	$\cdots$	$\cdots$	$\cdots$

* Padé, Thesis, published in the Annales de l'Ecole Normale Supérieure, (3), vol 9, supplement, pp. 1–93, 1892.

where $N_{m,n}(z)/D_{m,n}(z)$ is the rational fraction, taken in its *lowest terms*, in which the degrees of numerator and denominator do not exceed n and m respectively, and whose expansion into a power series, $P(z)$, agrees with (8) for a greater number of terms than any other such rational fraction.

If a particular approximant does not appear more than once in the table, that approximant is said to be normal, and when every approximant is normal, the table is called *normal*. The power series $P(z)$ for a normal approximant, in which necessarily the degrees of numerator and denominator are exactly n and m respectively, agrees with the given series term by term up to and including the term of degree $m+n$ and no further.

Professor Van Vleck* showed that the series of Stieltjes (5) has a normal Padé table. The odd convergents of the continued fraction (7) fill the principal diagonal of the table, while the even convergents fill the parallel diagonal file immediately below it. Consequently these two diagonal files of approximants will either converge to a common limit analytic throughout the entire plane except over the whole or a part of the negative real axis, or to two distinct functions, meromorphic everywhere except at the origin, according as $\sum a_i$ diverges or converges.

The first main question which I have considered in this article is the following:

1. *What holds regarding the convergence of the other diagonal files parallel to the principal diagonal in the Padé table for the series of Stieltjes?*

I find that the sequence of convergents in any file parallel to the principal diagonal converges to a limit. Three cases arise.

Case I. If the series $\sum a_i$ is convergent, no two files to the right of and parallel to the principal diagonal have the same limit, and no two parallel files below the principal diagonal have the same limit. In this case all the limit functions are meromorphic over the entire plane except at the origin. The poles of the limits of the files to the right of the principal diagonal are all simple and lie upon the negative half of the real axis.

When $\sum a_i$ diverges there are two possibilities which exist, namely:

Case II. The file of approximants in the principal diagonal and the successive parallel files to the right of it up to the nth file inclusive converge to one and the same limit, while subsequent to the nth they converge to distinct limits. Likewise the successive parallel files below the principal diagonal, down to an mth file inclusive, may converge to the same limit, while subsequent to the mth file they converge to distinct limits.

* *On an extension of the* 1894 *memoir of Stieltjes*, these Transactions, vol. 4 (1903), pp. 297–332.

All the limit functions in this case are meromorphic over the entire plane except at the origin.

Case III. Every diagonal file parallel to the principal diagonal converges to one and the same function, which is analytic over the entire plane excepting the whole or a part of the negative half of the real axis.

In Case I, the series (5) diverges, while in Cases II and III it may either diverge or converge. Whereas Stieltjes obtained two distinct functional equivalents of the divergent series (5) in Case I, we now have an infinite set; and whereas Stieltjes obtained a single functional equivalent when $\sum a_i$ diverges, I obtain an infinite number in Case II, and but one in Case III.

Generalizing (1), I shall call any series

$$\pm(c_0 z^n - c_1 z^{n+1} + c_2 z^{n+2} - \cdots), \tag{9}$$

in which n is a positive or negative integer or zero, a Stieltjes series if the coefficients c_i satisfy the Stieltjes inequalities (6).

If the first n terms of (5) be removed, the remainder is still a Stieltjes series. Instead of removing terms I have concerned myself with the following question:

2. *Can the series* (5) *be so continued to the left by successive addition of new terms that the extended series shall be Stieltjes series?*

I find that the necessary and sufficient condition for a first extension is that the series $\sum a_{2i}$, with even subscripts, shall converge.

Two cases arise:

Case A. If the series $\sum a_i$ converges, the Stieltjes series (5) may be extended to the left by the addition of an arbitrary number of new terms, forming thereby a series whose coefficients satisfy the Stieltjes inequalities (6). The coefficients of the prefixed terms are not unique.

Case B. If the series $\sum a_i$ diverges and $\sum a_{2i}$ converges, the series (5) may be extended by the addition of a term, $-c_{-1}z^{-1}$, in which c_{-1} may be taken equal to or greater than $\sum a_{2i}$. If $c_{-1} > \sum a_{2i}$, further extension is never possible; but if $c_{-1} = \sum a_{2i}$, the like series $\sum a'_{2i}$ in the new Stieltjes continued fraction belonging to the extended series may or may not converge. If it does converge, a second extension is possible. The coefficient of the new term, $c_{-2}z^{-2}$, prefixed may be taken equal to or greater than $\sum a'_{2i}$. Only if $c_{-2} = \sum a'_{2i}$, can a third term be prefixed, and so on.

When n terms can be prefixed, all the coefficients in the added terms are unique with the exception of c_{-n}. For an infinite extension all the coefficients are unique. There exist series which can be extended to just n terms and others which can be infinitely extended.

CHAPTER I. THE PADÉ TABLE OF APPROXIMANTS

To each pair of non-negative integers (m, n) there corresponds one rational function of the complex variable $z=x+iy$: $N_{m,n}(z)/D_{m,n}(z)$ taken in its *lowest* terms, in which the degrees of numerator and denominator do not exceed n and m respectively, and whose expansion into a series in ascending powers of z agrees term by term with a given series

$$k_0 + k_1 z + k_2 z^2 + \cdots \qquad (k_0 \neq 0) \tag{8}$$

to a higher degree than all other such rational functions. In fact, if we let $N_{m,n}(z)=s_0+s_1z+\cdots+s_nz^n$ and $D_{m,n}(z)=t_0+t_1z+\cdots+t_mz^m$, the conditions that the series

$$(k_0 + k_1 z + \cdots)(t_0 + t_1 z + \cdots + t_m z^m) - (s_0 + s_1 z + \cdots + s_n z^n) \tag{10}$$

shall begin with the $(m+n)$th power of z are the following:

$$k_i t_0 + k_{i-1} t_1 + \cdots + k_{i-m} t_m = s_i \qquad (i = 0,1, \cdots, n;\ k_j = 0 \text{ if } j < 0), \tag{11}$$

$$k_{n+i} t_0 + k_{n+i-1} t_1 + \cdots + k_{n+i-m} t_m = 0 \qquad (i = 1,2, \cdots, m;\ k_j = 0 \text{ if } j < 0). \tag{12}$$

If we take t_0 equal to the determinant

$$\Delta_{m-1,n} \equiv \begin{vmatrix} k_{n-m+1} & \cdots & k_n \\ \cdot & & \cdot \\ k_n & \cdots & k_{n+m-1} \end{vmatrix},$$

and assume $\Delta_{m-1,n} \neq 0$, then all the other t_i and also the s_i are determined uniquely, and $N_{m,n}$ and $D_{m,n}$ may be written in the form of determinants:

$$N_{m,n}(z) = \begin{vmatrix} k_{n-m+1} & \cdots & k_n & k_{n-m}z^n + k_{n-m-1}z^{n-1} + \cdots + k_0 z^m \\ \cdot & & \cdot & \cdot \\ k_{n+1} & \cdots & k_{n+m} & k_n z^n + k_{n-1} z^{n-1} + \cdots + k_0 \end{vmatrix}, \tag{13}$$

$$D_{m,n}(z) = \begin{vmatrix} k_{n-m+1} & \cdots & k_n & z^m \\ \cdot & & \cdot & \cdot \\ k_{n+1} & \cdots & k_{n+m} & 1 \end{vmatrix}. \tag{14}$$

The table of approximants is normal when and only when all the determinants

$$\Delta_{m,n} \equiv \begin{vmatrix} k_{n-m} & \cdots & k_n \\ \cdot & & \cdot \\ k_n & \cdots & k_{n+m} \end{vmatrix} \qquad \begin{pmatrix} m, n = 0, 1, 2, \cdots \\ k_i = 0, \text{ for } i < 0 \end{pmatrix}$$

are different from zero.* We will henceforth assume the normality of the table.

In a normal table Padé distinguished three types of regular continued fractions. We shall need to use here the *stairlike* type. This is obtained by starting from a border element of the table of approximants, and passing hence alternately one step to the right and one downward or vice versa. If, for example, we start from the nth element of the upper border of the table and proceed hence one step to the right and one down alternately, we get the sequence of approximants

$$\frac{N_{0,n-1}(z)}{D_{0,n-1}(z)}, \quad \frac{N_{0,n}(z)}{D_{0,n}(z)}, \quad \frac{N_{1,n}(z)}{D_{1,n}(z)}, \quad \frac{N_{1,n+1}(z)}{D_{1,n+1}(z)}, \cdots \tag{15}$$

whose terms are the successive convergents of the continued fraction

$$k_0 + k_1 z + k_2 z^2 + \cdots + k_{n-1} z^{n-1} + z^n \left[\frac{1}{h_1^{(n)}} + \frac{z}{h_2^{(n)}} + \frac{z}{h_3^{(n)}} + \cdots \right], \tag{16}$$

in which

$$h_{2i}^{(n)} = -\frac{\Delta_{i-1,i+n-1}^2}{\Delta_{i-1,i+n}\Delta_{i-2,i+n-1}}, \qquad h_{2i+1}^{(n)} = \frac{\Delta_{i-1,i+n}^2}{\Delta_{i-1,i+n-1}\Delta_{i,i+n}}. \tag{17}$$

If we start with the corner element of the table and proceed hence one step downward and one to the right alternately, we get the sequence of approximants

$$\frac{N_{0,0}(z)}{D_{0,0}(z)}, \quad \frac{N_{1,0}(z)}{D_{1,0}(z)}, \quad \frac{N_{1,1}(z)}{D_{1,1}(z)}, \quad \frac{N_{2,1}(z)}{D_{2,1}(z)}, \cdots, \tag{18}$$

whose terms are the successive convergents of a continued fraction

$$\frac{1}{a_1} + \frac{z}{a_2} + \frac{z}{a_3} + \frac{z}{a_4} + \cdots, \tag{19}$$

in which

$$a_{2i} = -\frac{\Delta_{i-1,i-1}^2}{\Delta_{i-1,i}\Delta_{i-2,i-1}}, \qquad a_{2i+1} = \frac{\Delta_{i-1,i}^2}{\Delta_{i-1,i-1}\Delta_{i,i}}.$$

It is to be understood that $\Delta_{-1,0}$ is to be taken here and hereafter equal to unity.

If now the continued fraction (16) converges, the two adjacent diagonal files of the table from which (15) is obtained by selecting approximants from each alternately, converge to one and the same limit, namely the limit of (16). On the other hand, if these two diagonal files converge to separate

* For details, see Padé, Thesis, loc. cit; or O. Perron, *Die Lehre von den Kettenbrüchen*, 1913, Chap. X, pp. 418–465.

limits, the sequences of alternate convergents of (16) must have separate limits.

It is easily seen that not only does the series (8) uniquely determine the continued fraction (19) in the case of a normal table but conversely (19) determines the series. The study of the convergence of the files of approximants parallel to the principal diagonal of the table can therefore be made to depend upon the sequence of numbers a_i in (19). In the next chapter such a study is made for the diagonal files to the right of the principal diagonal of the table in the case of the continued fraction of Stieltjes.

To investigate the convergence of the diagonal files below the principal diagonal which start on the left border of the table, I find it desirable to turn to the Padé table of the reciprocal series

$$\delta_0 + \delta_1 z + \delta_2 z^2 + \delta_3 z^3 + \cdots, \tag{20}$$

which is determined by the identity

$$(k_0 + k_1 z + \cdots)(\delta_0 + \delta_1 z + \cdots) = 1. \tag{21}$$

The approximant corresponding to (m, n) for this series is the reciprocal of the approximant corresponding to (n, m) for the original series (8); and when, as supposed, the Padé table of (8) is normal, the table of the reciprocal series is also normal, inasmuch as the first $m+n+1$ terms of the latter series are determined by the first $m+n+1$ terms of the former and conversely. Thus the question of the convergence of the diagonal files to the left of the principal diagonal in a normal Padé table may be replaced by the consideration of the convergence of the diagonal files to the right of the principal diagonal in the table of the reciprocal series.

Let now $N'_{n,m}(z)/D'_{n,m}(z)$, $m>n$, be an approximant of the series (20). Then

$$D'_{n,m} = w(m, n-m)\, N_{m,n}, \tag{22}$$

$$N'_{n,m} = w(m, n-m) D_{m,n}, \tag{23}$$

where $w(m, n-m)$ is a numerical factor independent of z. By (13), (14) the relation (22) may be written

$$\begin{vmatrix} \delta_{n-m+1} & \cdots & \delta_n, & z^n \\ \cdot & \cdot & \cdot & \cdot \\ \delta_{n+1} & \cdots & \delta_{n+m}, & z^0 \equiv 1 \end{vmatrix} = w(m, n-m) \begin{vmatrix} k_{m-n+1} & \cdots & k_m, & k_{m-n} z^m + k_{m-n-1} z^{m-1} + \cdots \\ \cdot & \cdot & \cdot & \cdot \\ k_{m+1} & \cdots & k_{m+n}, & k_m z^m + k_{m-1} z^{m-1} + \cdots \end{vmatrix}, \tag{24}$$

with a similar equation resulting from (23).

On setting $z=0$, and replacing n by $n+m$ in (24) the latter becomes

$$\Delta'_{m-1,n+m} = k_0 w(m,n)\cdot\Delta_{m+n-1,m}, \tag{25}$$

where $\Delta'_{m,n}$ denotes the determinant $\Delta_{m,n}$ expressed in terms of the δ_i instead of the k_i.

We will now determine the constant $w(m, n)$.

From the identity (21) we have

$$\begin{aligned} \delta_0 k_0 &= 1, \\ \delta_1 k_0 + \delta_0 k_1 &= 0, \\ &\cdots \\ \delta_n k_0 + \delta_{n-1} k_1 + \cdots + \delta_0 k_n &= 0, \end{aligned} \tag{26}$$

which gives

$$\delta_n = (-1)^{n(n+1)/2}\cdot\frac{1}{k_0^{n+1}}\cdot\Delta_{n-1,1}. \tag{27}$$

Also, setting $m=1$ and replacing n by $n-1$ in (25) we have

$$\delta_n = k_0 w(1, n-1)\Delta_{n-1,1}, \tag{28}$$

which with (27) gives

$$w(1,n) = (-1)^{(n+1)(n+2)/2}\cdot\frac{1}{k_0^{n+3}}. \tag{29}$$

I shall now prove by mathematical induction that*

$$w(m,n) = (-1)^{(n+1)(n+2)/2+m-1}\cdot\frac{1}{k_0^{2m+n+1}}. \tag{30}$$

We see from (29) that the value of $w(m, n)$ given by (30) is correct for all values of n if $m=1$.

I shall assume the correctness of formula (30) for all values of n when $m=1, 2, 3, \cdots, p-1$, and prove, on that hypothesis, that it is correct for $m=p$, and hence for all m.

By a well known rule for multiplying a determinant by one of its minors,

$$\Delta'_{p-1,n+p} = \frac{\Delta'_{p-2,n+p-1}\Delta'_{p-2,n+p+1} - (\Delta'_{p-2,n+p})^2}{\Delta'_{p-3,n+p}}. \tag{31}$$

Then by (25) and what we have assumed, (31) may be written

$$\Delta'_{p-1,n+p} = \frac{(-1)^{(n+1)(n+2)/2+p-1}\left[\dfrac{-\Delta_{n+p,p-1}}{\Delta_{n+p-1,p}} - \dfrac{\Delta^2_{n+p-1,p-1}}{\Delta_{n+p-1,p}\Delta_{n+p-2,p-1}}\right]}{-k_0^{2p+n+2}(\Delta_{n+p-1,p-2}/\Delta_{n+p-1,p}\Delta_{n+p-2,p-1})}. \tag{32}$$

* Cf. Hadamard, Liouville's Journal, (4), vol. 8 (1892), §24, p. 138. Hadamard obtains (25) with $w(m, n)$ as in (30) but by a different method.

But by the above mentioned rule for multiplying a determinant by one of its minors, the quantity within the bracket is equal to $-\Delta_{n+p-1,p-2}/\Delta_{n+p-2,p-1}$. Consequently, (32) may be written

$$\Delta'_{p-1,n+p} = \frac{(-1)^{(n+1)(n+2)/2+p-1}\Delta_{n+p-1,p}}{k_0^{2p+n+2}}. \tag{33}$$

Comparison of (25) and (33) shows that the value of $w(m, n)$ as given by (30) is correct for $m=p$, as was to be proved.

As particular instances of (25) we have

$$\Delta'_{m-1,m} = (-1)^m\Delta_{m-1,m}/k_0^{2m}, \tag{34}$$

$$\Delta'_{m-1,m-1} = (-1)^{m-1}\Delta_{m-2,m}/k_0^{2m-1}. \tag{35}$$

Due to symmetry of the identity (21), the δ_i and the k_i may be interchanged in (35), giving the relation

$$\Delta'_{m-2,m} = (-1)^{m-1}\Delta_{m-1,m-1}/k_0^{2m-1}. \tag{36}$$

We will now consider the continued fraction of the form (19) belonging to the normal* series

$$\delta_1 + \delta_2 z + \delta_3 z^2 + \cdots. \tag{37}$$

If we denote by a'_i the a_i obtained from (37), then by (34), (35), and (36) we obtain

$$a'_{2i} = -(\Delta'_{i-1,i})^2/[\Delta'_{i-1,i+1}\Delta'_{i-2,i}] = \Delta^2_{i-1,i}/[\Delta_{i,i}\Delta_{i-1,i-1}] = a_{2i+1},$$

$$a'_{2i+1} = (\Delta'_{i-1,i+1})^2/[\Delta'_{i-1,i}\Delta'_{i,i+1}] = -\Delta^2_{i,i}/[\Delta_{i-1,i}\Delta_{i,i+1}] = a_{2i+2}.$$

We have thus proved the following theorem:

THEOREM 1. *If $k_0+k_1z+k_2z^2+\cdots$ is any normal power series giving rise to the continued fraction*

$$\frac{1}{a_1} + \frac{z}{a_2} + \frac{z}{a_3} + \cdots \tag{38}$$

and $\delta_0+\delta_1z+\delta_2z^2+\cdots$ its reciprocal, then the continued fraction of the form (19) *associated with the series $\delta_1+\delta_2z+\delta_3z^2+\cdots$ is*

$$\frac{1}{a_2} + \frac{z}{a_3} + \frac{z}{a_4} + \cdots. \tag{39}$$

* A normal series is one giving rise to a normal Padé table.

CHAPTER II. THE CONVERGENCE OF THE DIAGONAL FILES FOR THE SERIES OF STIELTJES

I pass now to the series of Stieltjes

$$c_0 - c_1 z + c_2 z^2 - \cdots . \tag{5}$$

The work of the preceding chapter is applicable here if we replace throughout k_{2i} by c_{2i} and k_{2i-1} by $-c_{2i-1}$. In particular the continued fraction (16) becomes*

$$G_n(z) + (-z)^n \left[\frac{1}{a_1^n} + \frac{z}{a_2^n} + \frac{z}{a_3^n} + \cdots \right], \tag{40}$$

where $G_n(z) = c_0 - c_1 z + \cdots + (-1)^{n-1} c_{n-1} z^{n-1}$ and the a_i^n are obtained by replacing every k_{i+n} by $(-1)^i c_{i+n}$ in (17). It is easily seen that now, in place of (17), we have

$$\begin{aligned} a_{2i}^n &= \Delta_{i-1,i+n-1}^2 / [\Delta_{i-1,i+n} \Delta_{i-2,i+n-1}], \\ a_{2i+1}^n &= \Delta_{i-1,i+n}^2 / [\Delta_{i-1,i+n-1} \Delta_{i,i+n}], \end{aligned} \tag{41}$$

in which it is to be understood that the $\Delta_{m,n}$ are the same as in Chapter I with every k_i replaced by c_i.

In particular,

$$\begin{aligned} a_{2i} &= A_i^2 / (B_i B_{i-1}) = \Delta_{i-1,i-1}^2 / (\Delta_{i-1,i} \Delta_{i-2,i-1}), \\ a_{2i+1} &= B_i^2 / (A_i A_{i+1}) = \Delta_{i-1,i}^2 / (\Delta_{i-1,i-1} \Delta_{i,i}). \end{aligned} \tag{42}$$

Thus each a_i^n is obtained from the corresponding a_i by means of the substitution of c_{i+n} for c_i throughout.

Since, by hypothesis, A_n and B_n, for $n = 1, 2, 3, \cdots$, are positive, all the determinants $\Delta_{m,n}$ formed from them by removing the first r rows and r columns are positive.† Therefore, in (40), the bracketed part,

$$\frac{1}{a_1^n} + \frac{z}{a_2^n} + \frac{z}{a_3^n} + \cdots , \tag{43}$$

is a continued fraction of Stieltjes with positive a_i^n. We may therefore apply his convergence results for the continued fraction (2), bearing in mind that, as noted in the introduction, z has been replaced by $1/z$ in (2) and then the factor z has been dropped. It is readily seen that if $P_i(z)$, $Q_i(z)$ denote the numerator and denominator of the convergents of (2) (after a_i has been re-

* Henceforth for convenience I write a_i^n in place of $a_i^{(n)}$.

† Sylvester, Philosophical Magazine, (4), vol. 4 (1852), pp. 140–141.

placed by a_i^n), and if $P_i'(z)$, $Q_i'(z)$ denote the corresponding numerators and denominators of (43), then*

$$P_{2i}'(z) = z^{i-1}P_{2i}\left(\frac{1}{z}\right), \qquad Q_{2i}'(z) = z^iQ_{2i}\left(\frac{1}{z}\right),$$

$$P_{2i+1}'(z) = z^iP_{2i+1}\left(\frac{1}{z}\right), \quad Q_{2i+1}'(z) = z^{i+1}Q_{2i+1}\left(\frac{1}{z}\right).$$

The sequence of convergents accordingly separates into two sets, the even and the odd convergents. Stieltjes showed that when $\sum a_i^n$ is convergent, the numerators and denominators of the even convergents converge to limits $p(z)$, $q(z)$ respectively; likewise those of the odd convergents to limits $p_1(z)$, $q_1(z)$, and these limits are *entire* functions whose zeros all lie upon the negative half of the real axis. These functions are connected by the relation

$$p_1(z)q(z) - p(z)q_1(z) = 1. \tag{44}$$

On the other hand, when $\sum a_i^n$ is divergent, the even and the odd convergents converge to a common function $F(z)$ which is analytic over the entire plane excepting the whole or a part of the negative half of the real axis.

If now we denote the ith convergent of (40) by $U_i(z)/V_i(z)$, we have

$$\frac{U_i(z)}{V_i(z)} = G_n(z) + (-z)^n\frac{P_i'(z)}{Q_i'(z)} = G_n(z) - (-z)^{n-1}\frac{P_i\left(\frac{1}{z}\right)}{Q_i\left(\frac{1}{z}\right)}. \tag{45}$$

The odd and even convergents of (40) fill respectively the nth and $(n+1)$st diagonal files of the Padé table. When $\sum a_i^n$ is convergent, they have separate limits in consequence of the foregoing results of Stieltjes, and these limits are meromorphic over the entire plane except at $z=0$, with poles lying only upon the negative half of the real axis. If now we put

$$u(z) = G_n(z)q\left(\frac{1}{z}\right) - (-z)^{n-1}p\left(\frac{1}{z}\right),$$

$$v(z) = q\left(\frac{1}{z}\right),$$

the limit of the even convergents $U_{2i}(z)/V_{2i}(z)$, for $i=\infty$, is $u(z)/v(z)$. Similarly the limit of the odd convergents will be $u_1(z)/v_1(z)$ where u_1 and v_1

* Cf. the formulas for $P_i(z)$, $Q_i(z)$ given by Stieltjes, loc. cit., §2.

are the same as u and v except that p, q are replaced by p_1, q_1. In consequence of (44) we have the relation

$$u(z)v_1(z) - u_1(z)v(z) = (-z)^{n-1}. \tag{46}$$

When $\sum a_i^n$ is divergent, the application of the results of Stieltjes to (45) shows immediately that the two sets of convergents have a common limit which is analytic over the entire plane except the whole or a part of the negative real axis. We conclude therefore that whether $\sum a_i^n$ converges or diverges every diagonal file of approximants to the right of the principal diagonal converges.

For a complete discussion of the relationship between the limits of these files we shall need to connect our successive series $\sum a_i^n$ with the primary series $\sum a_i$. For that purpose I shall now obtain expressions for the a_i^n in terms of the a_i^{n-1}.

Let the numerator and denominator of the ith convergent of (40) be denoted by $U_i^n(z)$ and $V_i^n(z)$. Except for a common numerical factor, these may be obtained from (13) and (14). Thus, in particular, on setting $m=k$ and replacing n by $n+k$ in (14) and introducing this factor, we have*

$$V_{2k+1}^n(-z) = (-1)^k \begin{vmatrix} c_{n+1} & \cdots & c_{n+k} & z^k \\ \cdot & \cdot & \cdot & \cdot \\ c_{n+k+1} & \cdots & c_{n+2k} & 1 \end{vmatrix} : \Delta_{k,n+k}. \tag{47}$$

From (47) we see that the coefficient of z^k in $V_{2k+1}(z)$ is $\Delta_{k-1,n+k+1}/\Delta_{k,n+k}$. Now $V_{2k+1}^n(z)$ is the denominator of the $(2k+1)$st convergent of the Stieltjes continued fraction (43), and it is easily proved by mathematical induction that the coefficient of z^k is also $a_1^n + a_3^n + \cdots + a_{2k+1}^n$. Therefore

$$a_1^n + a_3^n + \cdots + a_{2k+1}^n = \Delta_{k-1,n+k+1}/\Delta_{k,n+k}. \tag{48}$$

If now we form by (41) the ratios a_{2i}^n/a_{2i+1}^{n-1} and $a_{2i+1}^n/a_{2i+2}^{n-1}$ and then employ (48), we obtain the very fundamental relations†

$$a_{2i}^n = \frac{a_{2i+1}^{n-1}}{\sum_{i=0}^{i-1} a_{2i+1}^{n-1} \sum_{i=0}^{i} a_{2i+1}^{n-1}} \qquad (n = 1,2,3,\cdots), \tag{49}$$

* Cf. Perron, loc. cit., p. 428. Note that Perron's $\Delta_{\mu\nu}$ differs by a factor $(-1)^{\mu(\mu+1)/2}$ from that used by me.

† After this article was finished I found these relations (unused) in Stieltjes, loc. cit., §78.

and

$$a_{2i+1}^{n} = a_{2i+2}^{n-1}\left(\sum_{i=0}^{i} a_{2i+1}^{n-1}\right)^2. \tag{50}$$

For $n=0$ we are to understand that $a_i^0 \equiv a_i$ $(i=1, 2, 3, \cdots)$.

Equation (49) may evidently be written in the form

$$a_{2i}^{n} = \frac{1}{\sum_{i=0}^{i-1} a_{2i+1}^{n-1}} - \frac{1}{\sum_{i=0}^{i} a_{2i+1}^{n-1}}. \tag{49a}$$

It now follows from (49) that when $\sum a_{2i-1}^{n-1}$ is convergent and has the limit L, then from and after some index i,

$$a_{2i}^{n} < a_{2i+1}^{n-1}/(L-e)^2,$$

where e is an arbitrary fixed positive number. Therefore $\sum a_{2i}^{n}$ is convergent when $\sum a_{2i-1}^{n}$ is convergent. Furthermore, $\sum a_{2i}^{n}$ is convergent even when $\sum a_{i-1}^{n}$ is divergent. For by summing (49a) for $i=1, 2, 3, \cdots, k$, we get

$$\sum_{i=1}^{k} a_{2i}^{n} = \frac{1}{a_1^{n-1}} - \frac{1}{\sum_{i=0}^{k} a_{2i+1}^{n-1}},$$

and this has the limit $1/a_1^{n-1}$ when k becomes infinite. We thus have proved the theorem:

THEOREM 2. *The successive series* $\sum a_{2i}^{n}$, $n=1, 2, 3, \cdots$, *are all convergent.*

We now turn to the series with odd subscripts, $\sum a_{2i-1}^{n}$. We see at once by (50) that, when $\sum a_i^{n-1}$ is convergent,

$$a_{2i+1}^{n} < a_{2i+2}^{n-1}(L')^2 \qquad (i = 0, 1, 2, \cdots),$$

where $L' = \sum a_{2i-1}^{n-1}$, and therefore $\sum a_i^{n}$ is convergent. Taking $n=1, 2, 3, \cdots$ in succession, we obtain at once the following theorem:

THEOREM 3. *If in the Stieltjes continued fraction* (7) *the series* $\sum a_i$ *is convergent, then all the associated series* $\sum a_i^{n}$, $n=1, 2, 3, \cdots$, *are convergent.*

Consider next the case in which $\sum a_i$ diverges. Then, while $\sum a_{2i}^{1}$ will necessarily converge, $\sum a_{2i-1}^{1}$ may either converge or diverge. On the former hypothesis all subsequent series $\sum a_i^{n}$, $n=2, 3, 4, \cdots$, must converge, by application of Theorem 3. If $\sum a_{2i-1}^{1}$ diverges, then as before $\sum a_{2i}^{2}$ must converge but $\sum a_{2i-1}^{2}$ may either converge or diverge, etc. Proceeding in this manner we perceive that in addition to the possibility of the convergence of

all $\sum a_i^n$, $n=0, 1, 2, \cdots$, Case I, there are just two other possibilities: either, Case II, all of the series $\sum a_i^n$, $n=0, 1, 2, \cdots$, will diverge; or else, Case III, the first r will diverge while all subsequent to the rth will converge.

It remains to see that both the last possibilities can be realized. For this purpose the following lemma will be established.

LEMMA 1. *Let*

$$b_1 + b_2 + b_3 + \cdots$$

be any divergent series in which there is a limit to the ratio b_i/b_{i-1}: $\lim b_i/b_{i-1} = s$. *Then if* $s \geqq 1$ *we have*

$$\lim_{i=\infty} \frac{b_1 + b_2 + \cdots + b_{i+k}}{b_1 + b_2 + \cdots + b_i} = s^k.$$

According to the hypothesis, we may take a fixed j sufficiently large to insure that b_{j+k+i}/b_{j+i} $(i \geqq 0)$ shall lie between the limits $s^k - e'/2$ and $s^k + e'/2$ where e' is an arbitrarily small prescribed positive quantity. Then

$$\frac{b_{j+k} + b_{j+k+1} + \cdots + b_{j+k+i}}{b_j + b_{j+1} + \cdots + b_{j+i}} = \frac{1 - \dfrac{b_1 + b_2 + \cdots + b_{j+k-1}}{b_1 + b_2 + \cdots + b_{j+k+i}}}{1 - \dfrac{b_1 + b_2 + \cdots + b_{j-1}}{b_1 + b_2 + \cdots + b_{j+i}}} \cdot \frac{b_1 + b_2 + \cdots + b_{j+k+i}}{b_1 + b_2 + \cdots + b_{j+i}}$$

will lie between the same limits inasmuch as the ratio of each term of the numerator of the left member to the corresponding term of the denominator lies between these limits. But since $\sum b_i = +\infty$, the first factor on the right by taking i sufficiently large, say for $i > N$, becomes $1/(1+e'')$, where $|e''| < e'$. Thus

$$\left| \frac{b_1 + b_2 + \cdots + b_{j+k+i}}{b_1 + b_2 + \cdots + b_{j+i}} - s^k \right| \leqq e'(s^k + 2) < \epsilon \quad \text{if} \quad e' < \frac{\epsilon}{s^k + 2}, \; i > N.$$

The lemma accordingly results.

Construct now a Stieltjes continued fraction for which $\lim (a_{2i+1}/a_{2i-1}) = s$, $\lim (a_{2i}/a_{2i-2}) = r$. Let $0 < r < 1$, $s > 1$. Then by the lemma and (49), (50) for $n=1$ we obtain

$$\lim_{i=\infty} (a_{2i+1}^1/a_{2i-1}^1) = rs^2, \quad \lim_{i=\infty} (a_{2i}^1/a_{2i-2}^1) = 1/s.$$

If we so choose s that $rs^2 < 1$, $1/s < 1$, then $\sum a_i$ will diverge while $\sum a_i^1$ will converge. But if $rs^2 > 1$, $1/s < 1$, $\sum a_i^1$ as well as $\sum a_i$ will diverge. In

the latter event, we may again apply the lemma and (49), (50) for $n=2$ and obtain

$$\lim_{i=\infty}(a_{2i+1}^{2}/a_{2i-1}^{2}) = r^2s^3, \quad \lim_{i=\infty}(a_{2i}^{2}/a_{2i-2}^{2}) = 1/(rs^2).$$

If we so choose s that $rs^2>1$, $r^2s^3<1$, the series $\sum a_i$, $\sum a_i^1$ will diverge, while $\sum a_i^2$ will converge. But if $r^2s^3>1$, $1/(rs^2)<1$, $\sum a_i$, $\sum a_i^1$, and $\sum a_i^2$ will all diverge, and we may again apply the lemma, etc.

Continuing in this manner we find that if $0<r<1$, $1/r^{(k-1)/k}<s<1/r^{k/(k+1)}$, all of the series $\sum a_i$, $\sum a_i^1$, $\cdots$, $\sum a_i^{k-1}$ will diverge, while the series $\sum a_i^k$, $\sum a_i^{k+1}$, $\cdots$, will converge. When $s\geqq 1/r$, all of the series $\sum a_i^n$, $n=0, 1, 2, 3, \cdots$, are divergent. Thus the two possibilities indicated above when $\sum a_i$ is divergent can be realized.

Whenever the Stieltjes series (5) is convergent, the series $c_n-c_{n+1}z+c_{n+2}z^2-\cdots$, $n=1, 2, 3, \cdots$, are convergent, and therefore the numbers $1/(a_i^n a_{i+1})$, $i=1, 2, 3, \cdots$, must be bounded for each $n=1, 2, 3, \cdots$.* Hence none of the associated series $\sum a_i^n$, $n=1, 2, 3, \cdots$, can be convergent.

These associated series may also diverge even when the Stieltjes series (5) is *divergent*. For consider the Stieltjes continued fraction for which $a_{2i-1}=1/(ir^i)$, $a_{2i}=r^i$, $0<r<1$. In this the numbers $1/(a_ia_{i+1})$ increase without limit, and $\lim(a_{2i+1}/a_{2i-1})=1/r$, $\lim(a_{2i+2}/a_{2i})=r$. Accordingly the series (5) and the associated series diverge.

We have seen that there are *three cases* which arise. Of these we will now consider the first.

Case I. The series $\sum a_i$ is convergent.

Then by Theorem 3 all the series $\sum a_i^n$, $n=1, 2, 3, \cdots$, converge. Hence by the discussion at the beginning of the chapter all the diagonal files of approximants which lie to the right of the principal diagonal have limits which are meromorphic over the entire plane excepting $z=0$; with poles lying only on the negative half of the real axis; and the limits for no two successive files are identical.

To show that *no* two of these diagonal files have the same limit, I will show that for two of the denominators v_n of these limits the roots of $v_n(1/z)$ (different from 0) nearest the origin are distinct.

Stieltjes found that the roots of $Q_{2k}(z)$ and $Q_{2k+1}(z)$ for (2) are all real, non-positive, distinct, and separate each other. The former is of degree k and has no root equal to zero, while the latter is of degree $k+1$ and has a zero root. The same is true of $Q_{2k}^n(z)$, and $Q_{2k+1}^n(z)$, obtained from these by

* Cf. Stieltjes, loc. cit., §10.

replacing a_i by a_i^n. If then we denote by $r_k^{(n)}$ and $r_k^{(n+1)}$ the non-zero root nearest the origin of $Q_{2k}^n(z)$ and $Q_{2k+1}^n(z)$, we have

$$r_k^{(n)} > r_k^{(n+1)}. \qquad (n = 1,2,3, \cdots ; \; k = 1,2,3, \cdots).$$

But $r_k^{(n+1)}$ is also the non-zero root nearest the origin of $Q_{2k}^{n+1}(z)$. If then we let $k \to \infty$, $r_k^{(n)}$ will approach a limit $R_n \neq 0$, and this limit is the non-zero root nearest the origin of the limit function $v_n(1/z)$.* We have

$$R_1 \geqq R_2 \geqq R_3 \geqq R_4 \geqq \cdots . \tag{51}$$

The equality sign can nowhere hold in (51), since, by (46), $v_n(1/z)$ and $v_{n+1}(1/z)$ cannot vanish for the same value of z not zero, and therefore $R_n \neq R_{n+1}$ for all values of n. The distinctness of the functions $v_n(1/z)$ therefore follows, and hence no two diagonal files can converge to the same limit.

Case II. All the series $\sum a_i^n$, $n = 1, 2, 3, \cdots$, diverge.

Any two consecutive diagonal files to the right of and parallel to the principal diagonal converge to the same limit, and hence all these files have a common limit. The limit is analytic throughout the complex plane excepting the whole or a part of the negative half of the real axis.

Case III. The series $\sum a_i^n$, $n = 1, 2, 3, \cdots, r$, diverge while $\sum a_i^{r+j}$, $j = 1, 2, 3, 4, \cdots$, converge.

By the first part of the hypothesis the principal diagonal and all parallel files on its right up to the $(r+1)$th diagonal file inclusive converge to one and the same limit, while by virtue of the second part of the hypothesis all the files beginning with the $(r+1)$th converge to limits which are meromorphic over the entire plane except at $z = 0$. Thus the limits of *all* these files are of the same meromorphic character, but the distinctness of the limits begins only on passing the $(r+1)$th file.

We will now turn our attention to the diagonal files below the principal diagonal which start with an approximant on the vertical border of the Padé table. Instead of considering directly the convergence of these diagonal files of approximants, we may consider instead the convergence of their reciprocals which are files of approximants to the right of the principal diagonal of the Padé table for the reciprocal series of (5). Let this be

$$d_0 + d_1 z + d_2 z^2 + \cdots . \tag{52}$$

By (34) and (35) the determinants A_n and B_n with every c_i replaced by $(-1)^i d_i$ are not all positive, but to the series

$$d_1 + d_2 z + d_3 z^2 + \cdots \tag{53}$$

* Stieltjes, loc. cit., §19.

belongs (Theorem 1, Chapter I) a Stieltjes continued fraction

$$\frac{1}{a_2}+\frac{z}{a_3}+\frac{z}{a_4}+\cdots. \tag{54}$$

Consequently (53) is a Stieltjes series and we may apply the results previously obtained for (7) and (43) using (53) as our primary series.

Three cases may be distinguished.

Case I. If the series $\sum a_i$ converges, the diagonal files of approximants for (53) to the right of the principal diagonal converge each to a different function, which is meromorphic except at $z=0$ with poles only on the negative half of the real axis. Consequently the diagonal files of approximants for (5) below the principal diagonal converge, each to a different function, which is meromorphic except at $z=0$ with *zeros* only on the negative half of the real axis.

If $\sum a_i$ is divergent, two cases may be distinguished corresponding to II and III above.

Case II. If all the diagonal files of approximants for (53) to the right of the principal diagonal have a common limit $F_1(z)$, then those below the principal diagonal for (5) have the limit $F(z)=1/(d_0+zF_1)$, which is the limit of the continued fraction (7), and is analytic except over all or a part of the negative real axis.

Case III. If the first r files of approximants of (53) to the right of the principal diagonal have the same limit, while after passing the rth file the limits are distinct, then in the table of approximants of (5) the files below the principal diagonal will have the same limit until after passing the $(r+1)$th file, whereupon they become distinct from one another. The limit functions are all meromorphic except at $z=0$ with *zeros* only on the negative half of the real axis.

One may ask the question whether, if the rth and $(r+1)$th diagonal files to the right of the principal diagonal have a common [different] limit [limits], a corresponding pair of adjacent files below the principal diagonal must have a common [different] limit [limits]. In Case I the answer is obviously in the affirmative, for then all the files to the right of and also those below the principal diagonal have distinct limits. In Case II when (5) converges, since its reciprocal also converges, both the series $\sum a_i^n$, $n=0$, 1, 2, 3, $\cdots$, and the series similarly associated with (54) (which we shall denote for future reference by $\sum \alpha_i^n$, $n=0, 1, 2, \cdots$, $\alpha_i^0 \equiv a_{i+1}$) must all diverge. Hence *all* the files have the same limit. Also for the divergent Stieltjes series for which $a_{2i}=r^i$, $a_{2i-1}=1/(ir^i)$, $0<r<1$ (Chapter II) it is easily verified that $\sum a_i^n, \sum \alpha_i^n$, $n=0, 1, 2, 3, \cdots$, all diverge. Finally, when

$\lim(a_{2i}/a_{2i-2})=r$, $\lim(a_{2i+1}/a_{2i-1})=s$, $0<r<1$, $s>1$, it may be easily seen that a_{2i+1}/a_{2i-1}^{n} and $\alpha_{2i+1}^{n+1}/\alpha_{2i-1}^{n+1}$ both have the limit $r^n s^{n+1}$ for $i=\infty$. Consequently (except possibly when $r^n s^{n+1}=1$) $\sum a_i^n$ and $\sum \alpha_i^{n+1}$ converge or diverge together. In all these examples pairs of adjacent files equally distant from the *first diagonal file below* the principal diagonal behave alike (i.e. the two files in each pair have a common limit or else they have distinct limits). In the following example pairs of adjacent files equidistant from the *principal diagonal* behave alike. Choose $a_{2i+3}=b_i/(a_2+a_4+\cdots+a_{2i+2})^2$, where $\sum b_i$ is any convergent positive-term series. Then $\sum \alpha_{2i+1}^1=\sum b_i$ converges and hence by Theorem 3 all $\sum \alpha_i^n$ converge. But if $\sum a_{2i}$ is divergent, $\sum a_i$, $\sum a_i^1$ are divergent, but $\sum a_i^2$ and all subsequent series are convergent.

Now we shall examine the zeros and poles of the approximants $N_{m,n}(z)/D_{m,n}(z)$ of the Stieltjes series (5). If $n \geqq m$, these by (45) are equal to $G_{2p+1}(z)-z^{2p}p_j(1/z)/Q_j(1/z)$, where $j=2m$ or $2m+1$ according as $n-m=2p$ or $2p+1$; and if $n<m$ their reciprocals, by the discussion centering about (52), are equal to $d_0+d_1z+\cdots+d_{2p+1}z^{2p+1}+z^{2p+1}P_j''(1/z)/Q_j''(1/z)$, where P_j''/Q_j'' is of the same form as P_j/Q_j and $j=2n$ or $2n+1$ as $m-n=2p+1$ or $2p+2$. If we replace z by $1/z$ and set $G_n^1(z)=d_0+d_1z+\cdots+d_{n-1}z^{n-1}$, these become

$$G_{2p+1}\left(\frac{1}{z}\right)-\frac{1}{z^{2p}}\frac{P_j(z)}{Q_j(z)} \quad \text{and} \quad G_{2p}'\left(\frac{1}{z}\right)+\frac{1}{z^{2p+1}}\frac{P_j''(z)}{Q_j''(z)},$$

where $P_j(z)/Q_j(z)$ and $P_j''(z)/Q_j''(z)$ are the jth convergents of Stieltjes continued fractions of the form (2). Now it follows from the work of Stieltjes that when $n \geqq m-1$ the roots of the polynomials $D_{m,n}(z)$ are all distinct and lie on the negative half of the real axis. If $n \leqq m$ the same is true of the roots of $N_{m,n}(z)$. If $n=m$ or $m-1$, all the roots of $N_{m,n}(z)$ and $D_{m,n}(z)$ are real and alternate along the negative half of the real axis. Consider now the roots of the numerator and denominator of an arbitrary approximant $N_{m,n}(z)/D_{m,n}(z)$. In the above representation for it, $P_j(z)/Q_j(z)$ and $P_j''(z)/Q_j''(z)$ are monotone functions of z which decrease from $+\infty$ to $-\infty$ as z increases between two adjacent roots of the denominators.* But the polynomials $G_n(1/z)$ and $G_n'(1/z)$ are continuous and bounded in any interval not containing the origin. Hence $G_{2p+1}(1/z)$ becomes equal to $P_j(z)/z^{2p}Q_j(z)$ at least once between every two adjacent non-zero roots of $Q_j(z)$. Therefore *if $n>m$, $N_{m,n}(z)$ vanishes at least once between every two adjacent roots of $D_{m,n}(z)$; similarly, if $n<m-1$, $D_{m,n}(z)$ vanishes at least once between every two adjacent roots of $N_{m,n}(z)$.*

* Cf. Stieltjes, loc. cit., §3.

CHAPTER III. EXTENDED STIELTJES SERIES

I pass now to the question whether, given the Stieltjes series (5), it is possible to find a number c_{-1} such that the series

$$-c_{-1}z^{-1}+c_0-c_1z+c_2z^2-\cdots \tag{55}$$

shall be a Stieltjes series, i.e. such that the determinants A_n and B_n in terms of the coefficients of (55) shall all be positive. The B_n for (55), being the A_n for the original series (5), are positive. Consequently it is both necessary and sufficient for the extension that a number $t=c_{-1}>0$ shall exist such that

$$f_i(t)=\begin{vmatrix} t & c_0 & \cdots & c_i \\ c_0 & c_1 & \cdots & c_{i+1} \\ \cdot & \cdot & & \cdot \\ c_i & & \cdots & c_{2i+1} \end{vmatrix}$$

shall be positive for all values of $i=0, 1, 2, 3, \cdots$.

Now

$$f_i(t)\equiv\Delta_{i,i+1}\cdot t+f_i(0). \tag{56}$$

Since the coefficient of t in (56) is positive, c_{-1} must be taken greater than the root of $f_i(t)$, i.e. greater than

$$t_i=-f_i(0)/\Delta_{i,i+1}, \tag{57}$$

for all values of i. It follows that the extension is possible when and only when the numbers t_i, $i=0, 1, 2, \cdots$, have an upper bound.

Now*

$$t_i=a_2+a_4+\cdots+a_{2i+2}. \tag{58}$$

In fact, by the rule for multiplying a determinant by one of its minors

$$f_i(t)\cdot\Delta_{i-1,i}=f_{i-1}(t)\Delta_{i,i+1}-\Delta_{i,i}^2. \tag{59}$$

If we divide both members of this equation by $\Delta_{i-1,i}\Delta_{i,i+1}$ and then replace i by $i-1, i-2, \cdots, 3, 2, 1, 0$, we have with the aid of (42)

$$\begin{aligned} a_{2i+2}&=f_{i-1}(t)/\Delta_{i-1,i}-f_i(t)/\Delta_{i,i+1}, \\ a_{2i}&=f_{i-2}(t)/\Delta_{i-2,i-1}-f_{i-1}(t)/\Delta_{i-1,i}, \\ &\cdot\qquad\qquad\cdot\qquad\qquad\cdot \\ a_4&=f_0(t)/\Delta_{0,1}-f_1(t)/\Delta_{1,2}. \end{aligned}$$

* Stieltjes, loc. cit., §35, eq. 11, obtained this in a different manner.

If to these equations we adjoin the identity

$$a_2 = t - f_0(t)/\Delta_{0,1},$$

and then add them together and set $t=t_i$, we obtain (58).

It follows from (58) that the numbers t_i, $i=0, 1, 2, \cdots$, will have an upper bound when and only when the series $\sum a_{2i}$ is convergent. When this condition is satisfied, we may take $c_{-1}=\sum a_{2i}+\sigma$ where σ is zero or an arbitrary positive number. We state this result in the following theorem.

THEOREM 4. *Given a Stieltjes series*

$$c_0 - c_1 z + c_2 z^2 - \cdots, \tag{5}$$

there exists a Stieltjes series

$$-c_{-1}z^{-1} + c_0 - c_1 z + c_2 z^2 - \cdots \tag{55}$$

when and only when the series $\sum a_{2i}$ in the Stieltjes continued fraction belonging to (5) *is convergent.*

When the condition is fulfilled the number c_{-1} may be taken equal to $\sum a_{2i}$ or any greater number.

Let us now suppose that $\sum a_i$ in (7) is convergent. Denote by*

$$\frac{1}{a_1^{-1}} + \frac{z}{a_2^{-1}} + \frac{z}{a_3^{-1}} + \cdots \tag{60}$$

the Stieltjes continued fraction belonging to (55), in which the a_n^{-1} are necessarily positive. We must now express the a_n^{-1} in terms of the a_n, which is obviously the same problem as that of expressing the a_i^{n-1} in terms of the a_i^n. To do this we will need to develop some necessary formulas.

If we equate the two values of t_{i-2} obtained from (57) and (58), and then advance the subscripts of every c_i by n, we get the equation

$$a_2^n + a_4^n + \cdots + a_{2i-2}^n = -f_{i-2}^n(0)/\Delta_{i-2,n+i-1}, \tag{61}$$

where

$$f_{i-2}^n(0) = \begin{vmatrix} 0 & c_n & \cdots & c_{n+i-2} \\ c_n & c_{n+1} & \cdots & c_{n+i-1} \\ \cdot & & \cdot & \cdot \\ c_{n+i-2} & & \cdots & c_{n+2i-3} \end{vmatrix}.$$

Now

$$\Delta_{i-1,n+i-2} = f_{i-2}^n(0) + c_{n-1}\Delta_{i-2,n+i-1}, \tag{62}$$

* The superscripts as before are written for conven ence without parentheses.

and therefore (61) may be written in the form

$$\Delta_{i-1,n+i-2} = f^n_{i-2}(0)\left(c_{n-1} - \sum_{i=1}^{i-1} a^n_{2i}\right). \tag{63}$$

By (41),

$$\frac{a^{n-1}_{2i}}{\Delta^2_{i-1,n+i-2}} = \frac{1}{\Delta_{i-1,i+n-1}\Delta_{i-2,i+n-2}}. \tag{64}$$

Squaring the members of (63) and multiplying them into the corresponding members of (64), we have

$$a^{n-1}_{2i} = a^n_{2i-1}\left(c_{n-1} - \sum_{i=1}^{i-1} a^n_{2i}\right)^2 \qquad (i = 2,3,4,\cdots); \tag{65}$$

while for $i=1$ we have

$$a_2^{n-1} = a_1^n c^2_{n-1}. \tag{66}$$

Also by (41), (63),

$$a^{n-1}_{2i+1} = \frac{a^n_{2i}}{\left(c_{n-1} - \sum_{i=1}^{i-1} a^n_{2i}\right)\left(c_{n-1} - \sum_{i=1}^{i} a^n_{2i}\right)} \qquad (i = 2,3,4,\cdots), \tag{67}$$

while for $i=0$, 1 we have

$$a_1^{n-1} = 1/c_{n-1}\,;\ a_3^{n-1} = a_2^n/[c_{n-1}(c_{n-1} - a_2^n)]. \tag{68}$$

Then we have, setting $n=0$, $c_{-1}=\sum a_{2i}+\sigma$, $\sigma>0$, in (65), (67),

$$a^{-1}_{2i} = a_{2i-1}\left(\sum a_{2i} + \sigma - \sum_{1}^{i-1} a_{2i}\right)^2 < a_{2i-1}\left(\sum a_{2i} + \sigma\right)^2,$$

and

$$a^{-1}_{2i+1} = a_{2i}\Big/\left(\sum a_{2i} + \sigma - \sum_{1}^{i-1} a_{2i}\right)\left(\sum a_{2i} + \sigma - \sum_{1}^{i} a_{2i}\right) < a_{2i}/\sigma^2.$$

Hence since $\sum a_i$ is convergent, it follows that $\sum a_i^{-1}$ is also convergent.

We may now take $c_{-2}=\sum a_{2i}^{-1}+\sigma'$, where $\sigma'>0$, and extend (55) to form the Stieltjes series

$$c_{-2}z^{-2} - c_{-1}z^{-1} + c_0 - c_1 z + c_2 z^2 - \cdots, \tag{69}$$

etc.

Continuing in this manner, we see that the series (5) can be extended indefinitely to the left. We have proved the following theorem:

THEOREM 5. *If in the Stieltjes continued fraction belonging to the Stieltjes series*

$$c_0 - c_1 z + c_2 z^2 - \cdots \tag{5}$$

the series $\sum a_i$ *converges, then the series* (5) *can be extended to the left by prefixing any finite or infinite number of terms:* $-c_{-1}z^{-1}, \quad +c_{-2}z^{-2}, \quad -c_{-3}z^{-3}, \cdots,$ *in such wise that the extended series shall be Stieltjes series.*

The coefficients in the prefixed terms are not unique.

Let us next suppose the series $\sum a_i$ to diverge. By Theorem 4, $\sum a_{2i}$ must converge in order that extension shall be possible, and consequently $\sum a_{2i-1}$ is divergent. We then take $c_{-1}=\sum a_{2i}+\sigma$, where σ is *zero* or a positive number. I shall prove that unless $\sigma=0$, the series $\sum a_{2i}^{-1}$ will be divergent, and therefore further extension will be impossible. In fact, by (65), if $\sigma \neq 0$, $a_{2i}^{-1} > \sigma^2 a_{2i-1}$, $i=1, 2, 3, \cdots$, and therefore each term of $\sum a_{2i}^{-1}$ is greater than the corresponding term of the divergent series $\sigma^2 \sum a_{2i-1}$ so that extension is impossible.

When $\sigma \neq 0$, it has been shown that $\sum a_{2i}^{-1}$ is necessarily divergent. I shall give examples presently to show that when $\sigma=0$, $\sum a_{2i}^{-1}$ may either converge or diverge. On the former hypothesis $\sum a_{2i-1}^{-1}$ must be divergent, and therefore also $\sum a_i^{-1}$. For suppose the contrary. Then by Theorem 3, since $\sum a_i^{-1}$ converges, $\sum a_i$, $\sum a_i^1$, $\cdots$ converge. But by hypothesis $\sum a_i$ is divergent. We thus have a contradiction. Hence:

THEOREM 6. *If in the Stieltjes continued fraction*

$$\cfrac{1}{a_1} + \cfrac{z}{a_2} + \cfrac{z}{a_3} + \cdots$$

belonging to the Stieltjes series

$$c_0 - c_1 z + c_2 z^2 - \cdots \tag{5}$$

the series $\sum a_i$ *diverges, and if* (5) *admits the extension*

$$\pm c_{-n} z^{-n} \mp c_{-n+1} z^{-n+1} \pm \cdots + c_0 - c_1 z + c_2 z^2 - \cdots,$$

then the coefficients $c_{-1}, c_{-2}, \cdots, c_{-n}$ *must be chosen successively in accordance with the relations*

$$\begin{aligned} c_{-i} &= \sum a_{2k}^{-i+1} \qquad (i = 1, 2, \cdots, n-1), \\ c_{-n} &\geqq \sum a_{2k}^{-n+1}. \end{aligned} \tag{70}$$

For examples illustrating the extension of Stieltjes series to the left the following lemma will be needed.

LEMMA 2. *Let* $b_1+b_2+b_3+\cdots$ *be any convergent series in which* $\lim_{i=\infty} b_i/b_{i-1}=r$; *then if* $0<r<1$,

$$\lim_{i=\infty} \frac{b_{i+1}+b_{i+2}+\cdots}{b_{i+k+1}+b_{i+k+2}+\cdots} = \frac{1}{r^k}.$$

According to the hypothesis we may take a fixed integer m sufficiently large to insure that for all $n \geqq m$ the quotient b_n/b_{n+k} will lie between $(1/r^k)-\epsilon/4$ and $(1/r^k)+\epsilon/4$, where ϵ is an arbitrarily small assigned positive number. Then

$$\frac{b_m+b_{m+1}+\cdots+b_{m+p}}{b_{m+k}+b_{m+k+1}+\cdots+b_{m+k+p}}$$

will lie between these same limits inasmuch as the terms are all of the same sign and the ratio of each term of the numerator to the corresponding term of the denominator lies between these limits. But since $\sum b_i$ is convergent there will be a limit to this fraction for $p\to\infty$ which lies between these limits.

As ϵ may be taken as small as desired by taking m sufficiently large, the lemma accordingly results.

Construct now a Stieltjes continued fraction in which $a_{2i}=r^i$, $\lim_{i=\infty} a_{2i+1}/a_{2i-1}=s$, $0<r<1$, $s>1$. As $\sum a_i$ diverges, c_{-1} must be taken in accordance with (70). Then by (65) and (67), with $n=0$,

$$\lim_{i=\infty} \frac{a_{2i}^{-1}}{a_{2i-2}^{-1}} = \lim_{i=\infty} \frac{a_{2i-1}}{a_{2i-3}} \frac{(a_{2i}+a_{2i+2}+\cdots)^2}{(a_{2i-2}+a_{2i}+\cdots)^2} = r^2s,$$

$$\lim_{i=\infty} \frac{a_{2i+1}^{-1}}{a_{2i-1}^{-1}} = \lim_{i=\infty} \frac{a_{2i}}{a_{2i-2}} \frac{(a_{2i-2}+a_{2i}+\cdots)}{(a_{2i+2}+a_{2i+4}+\cdots)} = 1/r.$$

If we now choose r so that $r^2s<1$, then $\sum a_{2i}^{-1}$ will be convergent and both a first and second extension of (5) is possible. This requires that we have $r<1/s^{1/2}<1$. Similarly, if $r^2s<1$, we have

$$\lim_{i=\infty} \frac{a_{2i}^{-2}}{a_{2i-2}^{-2}} = \lim_{i=\infty} \frac{a_{2i-1}^{-1}}{a_{2i-3}^{-1}} \frac{(a_{2i}^{-1}+\cdots)^2}{(a_{2i-2}^{-1}+\cdots)^2} = r^3s^2,$$

$$\lim_{i=\infty} \frac{a_{2i+1}^{-2}}{a_{2i-1}^{-2}} = \lim_{i=\infty} \frac{a_{2i}^{-1}}{a_{2i-2}^{-1}} \frac{(a_{2i-2}^{-1}+\cdots)}{(a_{2i+2}^{-1}+\cdots)} = 1/(r^2s).$$

Then if $r^3s^2<1$, that is, if $r<1/s^{2/3}$, a third extension is possible and $\lim a_{2i}^{-3}/a_{2i-2}^{-3}=r^4s^2$, etc. Thus, if

$$\frac{1}{s^{k/(k+1)}} < r < \frac{1}{s^{(k-1)/k}},$$

the series $\sum a_{2i}, \sum a_{2i}^{-1}, \cdots, \sum a_{2i}^{-k+1}$ will converge, while $\sum a_{2i}^{-k}$ is divergent. Hence extension of (5) to the left for k terms is then possible, but a further extension is impossible. If $r = 1/s$, an extension of infinite order is possible.

The following theorem follows immediately from (49), (50) of the preceding chapter.*

THEOREM 7. *Given any series of positive numbers $\sum a_i$, there is uniquely determined a power series of Stieltjes*

$$c_0 - c_1 z + c_2 z^2 - \cdots$$

belonging to the Stieltjes continued fraction

$$\frac{1}{a_1} + \frac{z}{a_2} + \frac{z}{a_3} + \cdots .$$

The c_i are given in terms of the a_i by the following formulas:

$$c_i = 1/a_1^i \qquad (i = 0,1,2, \cdots),$$

$$(71) \qquad a_{2k}^{i} = a_{2k+1}^{i-1} \Big/ \sum_{k=1}^{k} a_{2k-1}^{i-1} \sum_{k=1}^{k+1} a_{2k-1}^{i-1},$$

$$a_{2k-1}^{i} = a_{2k}^{i-1} \left(\sum_{k=1}^{k} a_{2k-1}^{i-1} \right)^2 .$$

CHAPTER IV. A PARTICULAR EXAMPLE

The continued fraction†

$$\frac{1}{1} + \frac{z}{1} + \frac{z}{1} + \frac{z}{1} + \cdots .$$

I will determine the c_i and the a_i^n (which are the same as the α_i^n in this case) for the above continued fraction. I will show that the radius of convergence of the corresponding Stieltjes series is 1/4. The associated series $\sum a_i^n$, $n = 1, 2, 3, \cdots$, are divergent, but the $\sum a_{2i}^n$ converge, and consequently

$$(72) \qquad c_n = \sum a_{2i}^{n+1} \qquad (n = 0,1,2,3, \cdots).$$

* Stieltjes, Annales de Toulouse, vol. 3, H, pp. 1–17, 1889, shows the one-to-one correspondence between his series and continued fractions. He obtains the c_i in terms of the a_i by a different method.

† Laurent, *Note sur les fractions continues*, Nouvelles Annales de Mathématiques, (2), vol. 5 (1866), pp. 540–552.

The c_n being known, (72) serve to give the sums of certain positive-term series.

To determine the c_i, write

$$F_1(z) = 1 + \cfrac{z}{1 + \cfrac{z}{1 + \cdots}} = 1 + \frac{z}{F_1(z)}.$$

Hence,

$$F_1(z) = \frac{1 + (1 + 4z)^{1/2}}{2},$$

and

$$\begin{aligned} F(z) &= \cfrac{1}{1 + \cfrac{z}{1 + \cfrac{z}{1 + \cdots}}} \\ &= (F_1 - 1)/z \\ &= \frac{-1 + (1 + 4z)^{1/2}}{2z}. \end{aligned}$$

If we develop the numerator of the last expression in a Taylor's power series about $z=0$, we may then divide by $2z$ and obtain

$$F(z) = c_0 - c_1 z + c_2 z^2 - \cdots, \tag{73}$$

where

$$c_{2k} = \frac{(2k+3)(2k+5)\cdots(4k-1)2^k}{k!}, \tag{74}$$

$$c_{2k+1} = \frac{(2k+3)(2k+5)\cdots(4k+1)2^k}{k+1!},$$

and therefore satisfy the recurrence relations

$$\begin{aligned} c_{2k+1} &= [(4k+1)/(k+1)]c_{2k}, \\ c_{2k} &= 2[(4k-1)/(2k+1)]c_{2k-1}, \qquad (75) \\ c_0 &= 1. \end{aligned}$$

Since $\lim(c_i/c_{i-1}) = 4$ it follows that the radius of convergence of (73) is equal to 1/4.

We will next get the elements in the associated continued fractions. Using (49), (50):

$$\begin{aligned} a^1_{2k} &= 1/[k(k+1)], \\ a^1_{2k+1} &= (k+1)^2. \end{aligned}$$

Then since $1^2+2^2+\cdots+(k+1)^2=(k+1)(k+2)(2k+3)/2\cdot 3$, we have

$$a_{2k}^{2} = (2\cdot 3)^2/[k(k+2)(2k+1)(2k+3)],$$
$$a_{2k+1}^{2} = (k+1)(k+2)(2k+3)^2/(2\cdot 3)^2.$$

The next step will involve the formula $1\cdot 2\cdot 3^2+2\cdot 3\cdot 5^2+\cdots+(k+1)\cdot(k+2)(2k+3)^2\equiv(k+1)(k+2)(k+3)(2k+3)(2k+5)/5$. We have

$$a_{2k}^{3} = (2\cdot 3)^2(5)^2/[k(k+1)(k+2)(k+3)(2k+1)(2k+5)],$$
$$a_{2k+1}^{3} = (k+1)(k+2)^2(k+3)(2k+3)(2k+5)/[(2\cdot 3)^2(5)^2].$$

In general:

$$a_{2k+1}^{2i+1} = \frac{(k+1)\cdots(k+i)(k+i+1)^2(k+i+2)\cdots(k+2i+1)(2k+3)\cdots(2k+4i+1)}{(2\cdot 3)^2(5)^2(2\cdot 7)^2\cdots[2(4i-1)]^2(4i+1)^2},$$

$$a_{2k}^{2i+1} = \frac{(2\cdot 3)^2(5)^2(2\cdot 7)^2\cdots[2(4i-1)]^2(4i+1)^2}{k\cdots(k+2i+1)(2k+1)\cdots(2k+2i-1)(2k+2i+3)\cdots(2k+4i+1)},$$

$$a_{2k+1}^{2i} = \frac{(k+1)\cdots(k+2i)(2k+3)\cdots(2k+2i-1)(2k+2i+1)^2(2k+2i+3)\cdots(2k+4i-1)}{(2\cdot 3)^2(5)^2(2\cdot 7)^2\cdots(4i-3)^2[2(4i-1)]^2},$$

$$a_{2k}^{2i} = \frac{(2\cdot 3)^2(5)^2(2\cdot 7)^2\cdots(4i-3)^2[2(4i-1)]^2}{k\cdots(k+i-1)(k+i+1)\cdots(k+2i)(2k+1)\cdots(2k+4i-1)}.$$

These formulas are seen to hold for all k when $i=1$. Assuming them true for all k and for all $i\leqq I-1$, they may, with the aid of the following formulas, be proved* for $i=I$. Set

$$Y_k^i = (k+1)\cdots(k+i-1)(k+i)^2(k+i+1)\cdots(k+2i-1)(2k+3)$$
$$\cdots(2k+4i-3),$$
$$Z_k^i = (k+1)\cdots(k+2i)(2k+3)\cdots(2k+2i-1)(2k+2i+1)^2(2k+2i+3)$$
$$\cdots(4k+4i-1).$$

Then,

$$\sum_{k=0}^{k} Y_k^i = (k+1)\cdots(k+2i)(2k+3)\cdots(2k+2i-1)(2k+2i+1)^2(2k+2i+3)$$
$$\cdots(2k+4i-1)/(2(4i-1)),$$

$$\sum_{k=0}^{k} Z_k^i = (k+1)\cdots(k+2i+1)(2k+3)\cdots(2k+4i+1)/(4i+1).$$

---

* The completion of the proof, which offers no difficulty, will be omitted.

UNIVERSITY OF WISCONSIN,
MADISON, WIS.

ON CURVILINEAR CONGRUENCES*

BY

C. E. WEATHERBURN

In a recent paper *On congruences of curves*† the writer has shown how the theory of such congruences in ordinary space of three dimensions may be extended along the lines followed for rectilinear congruences, making use of oblique curvilinear coördinates. The present paper contains a further extension of the theory, but is mainly independent of the preceding investigation. The subject is approached from a slightly different point of view, with the aid of the differential invariants grad, div, rot for three dimensions and, as far as possible, the use of coördinates is avoided. Some of the theorems proved have hitherto been known only for rectilinear congruences, while others again lose their significance in the particular case in which the lines are straight.

1. First quadric. Cone of zero tendency

Given a congruence of curves, the unit vector $\boldsymbol{t}$ tangent to the curve at any point P is known as a point function in space. It has a definite derivative for each direction. In the direction of the unit vector $\boldsymbol{a}$ its derivative‡ is $\boldsymbol{a}\cdot\nabla\boldsymbol{t}$, and the resolved part of this derivative in the direction of $\boldsymbol{a}$ has the value $\boldsymbol{a}\cdot\nabla\boldsymbol{t}\cdot\boldsymbol{a}$, the operator ∇ being understood to act only on the vector immediately following it. The quantity just defined may be called the *tendency* of the congruence at P in the direction of $\boldsymbol{a}$. It plays an important part in the following argument. Denoting it by T we have

$$T = \boldsymbol{a}\cdot\nabla\boldsymbol{t}\cdot\boldsymbol{a}. \tag{1}$$

If then we introduce the quadric§

$$\boldsymbol{r}\cdot\nabla\boldsymbol{t}\cdot\boldsymbol{r} = 1 \tag{2}$$

with a center at the point P, which is origin for the vector $\boldsymbol{r}$, it is clear that *the value of T for any direction at P is equal to the inverse square of the radius of the quadric* (2) *in that direction.* This square may be either positive or nega-

---

* Presented to the Society, October 29, 1927; received by the editors in September, 1927.

† *On congruences of curves*, Tôhoku Mathematical Journal, vol. 28 (1927), p. 114.

‡ Cf. the author's *Advanced Vector Analysis*, Art. 6.

§ Ibid., Art. 66.

tive. Also from the definition of the "divergence"* of a vector it follows that the sum of the tendencies for three mutually perpendicular directions at a point is invariant, and equal to div t. Hence:

The sum of the tendencies of the congruence in three mutually perpendicular directions at a point is invariant, and equal to the divergence of the congruence at that point.

The asymptotic cone of the quadric (2) is given by

$$\boldsymbol{r}\cdot\nabla\boldsymbol{t}\cdot\boldsymbol{r} = 0. \tag{3}$$

This may be called the *cone of zero tendency* at P, for in the direction of any of its generators, the tendency of the congruence is zero. The tangent to the curve at P is clearly a generator of this cone; for, since $\boldsymbol{t}$ is a unit vector, its derivative in the direction of the tangent is perpendicular to $\boldsymbol{t}$.

We shall be concerned largely with the section of the quadric (2) by the normal plane of the curve at P, that is to say, by the plane perpendicular to $\boldsymbol{t}$. This section is the conic

$$\boldsymbol{r}\cdot\boldsymbol{t} = 0, \qquad \boldsymbol{r}\cdot\nabla\boldsymbol{t}\cdot\boldsymbol{r} = 1 \tag{4}$$

whose asymptotes are the corresponding section of the cone (3), giving the directions of zero tendency in the normal plane. For a direction inclined at an angle θ in the normal plane to that of a principal axis of the conic (4), the tendency is given by

$$T = T_1\cos^2\theta + T_2\sin^2\theta$$

where T_1 is the tendency in the direction of the above axis, and T_2 that for the perpendicular direction. Obviously $T_1+T_2=\operatorname{div} t$, since the tendency in the direction of $\boldsymbol{t}$ is zero.

We may observe that a direction of zero tendency in the normal plane corresponds to that of a "common perpendicular" to nearby rays in the case of a rectilinear congruence.† This should be borne in mind in order to see the analogy which the following theorems present to those already known for a congruence of straight lines.

2. Surface of striction. Limit surface

The two directions of zero tendency in the normal plane will be at right angles provided div $\boldsymbol{t}$ is zero. This follows from the above theorem on the

---

* *Advanced Vector Analysis*, Art. 7.

† Cf. the author's *Differential Geometry*, Chap. X.

invariance of the sum of the tendencies in three perpendicular directions. The locus of the points at which this relation holds may be called the middle surface or *surface of striction* of the congruence, being analogous to the line of striction of a family of curves on a surface.* The points in which it is cut by any curve are the *points of striction* of the curve. Thus:

The locus of points at which the two directions of zero tendency in the normal plane are at right angles is the surface of striction, or middle surface, and is given by div $t=0$.

For a congruence of straight lines the middle surface† is the locus of points midway between the limits. In the case of a normal congruence the two directions of zero tendency in the normal plane are the asymptotic directions for the surface orthogonal to the curves. These directions are perpendicular when the mean curvature is zero. At such points div $t=0$ in agreement with the above.

The *limit surface*‡ may be defined as the locus of points at which the two directions of zero tendency in the normal plane are coincident. At such points the normal plane touches the cone of zero tendency. Now the normals to the quadric cone $\boldsymbol{r}\cdot\nabla t\cdot\boldsymbol{r}=0$ at its vertex generate another, called its reciprocal cone, whose equation is

$$\boldsymbol{r}\cdot(\nabla t)^{-1}\cdot\boldsymbol{r}=0$$

where $(\nabla t)^{-1}$ is the reciprocal dyadic to ∇t. Now the "second"§ of ∇t, which is denoted by $(\nabla t)_2$, is proportional to the conjugate of $(\nabla t)^{-1}$. Consequently the equation of the reciprocal cone may also be expressed as

$$\boldsymbol{r}\cdot(\nabla t)_2\cdot\boldsymbol{r}=0. \tag{5}$$

Thus, at points on the limit surface, the tangent to the curve must be a generator of the cone (5).

Let i, j, k be three fixed perpendicular unit vectors forming a right-handed system. Let i be parallel to t, and the other two therefore parallel to the normal plane of the curve at the point considered. Also let t_1, t_2, t_3 denote the derivatives of t in these three directions. Then

$$\nabla t = i\,t_1 + j\,t_2 + k\,t_3,$$

and

$$(\nabla t)_2 = i\,t_2\times t_3 + j\,t_3\times t_1 + k\,t_1\times t_2.$$

---

* *Differential Geometry*, Art. 126.

† *On congruences of curves*, Art 9.

‡ Ibid., Art. 4.

§ E. B. Wilson, *Vector Analysis*, pp. 316–317.

Hence the tangent to the curve will be a generator of the cone (5) provided $t\cdot(\nabla t)_2\cdot t=0$, that is

$$t\cdot(t_2\times t_3)=0. \tag{6}$$

Again, with the same notation, we have

$$t\,\mathrm{div}\,t-t\cdot\nabla t=i\times(t\times t_1)+j\times(t\times t_2)+k\times(t\times t_3)$$

from which it is easily verified that, t being a unit vector,

$$\mathrm{div}\,(t\,\mathrm{div}\,t-t\cdot\nabla t)=2\,i\cdot(t_2\times t_3)=2\,t\cdot(t_2\times t_3).$$

Thus the condition (6) is equivalent to

$$\mathrm{div}\,(t\,\mathrm{div}\,t-t\cdot\nabla t)=0.$$

Hence the theorem:

The locus of points at which the two directions of zero tendency in the normal plane are coincident is the limit surface, whose equation may be expressed as

$$\mathrm{div}\,(t\,\mathrm{div}\,t-t\cdot\nabla t)=0, \tag{7}$$

or

$$\mathrm{div}\,(t\,\mathrm{div}\,t+t\times\mathrm{rot}\,t)=0. \tag{7'}$$

In the case of a normal congruence of curves t is the unit normal to the surface orthogonal to the curves. Then, since the first member of (7) is twice the Gaussian curvature of the surface,* it follows that

The limit surface of a normal congruence is the locus of points at which the Gaussian curvature of the orthogonal surfaces is zero.

This agrees with the known property of a surface that the asymptotic directions are coincident where the Gaussian curvature vanishes.

3. Second quadric. Cone of zero moment

Again, let t be the unit tangent at P, and $t+\delta t$ that at a nearby point Q, such that the vector PQ is $\delta s\,a$, δs being the length of PQ and a a unit vector. The mutual moment of the tangents at P and Q, being the resolved part in the direction of t of the moment of $t+\delta t$ about P, has the value

$$\delta s\,a\times(t+\delta t)\cdot t=\delta s(a\times\delta t)\cdot t.$$

The quotient of this mutual moment by $(\delta s)^2$ has the value $a\times(\delta t/\delta s)\cdot t$,

* *Differential Geometry*, Art. 131.

and the limit of this as the point Q tends to coincidence with P, while the direction of $\boldsymbol{a}$ remains constant, is the function

$$M = \boldsymbol{a} \times (\boldsymbol{a}\cdot\nabla\boldsymbol{t})\cdot\boldsymbol{t} = \boldsymbol{a}\cdot(\nabla\boldsymbol{t} \times \boldsymbol{t})\cdot\boldsymbol{a}. \tag{8}$$

We shall call this the *moment* of the congruence at P for the direction of $\boldsymbol{a}$. Let us now introduce the quadric

$$\boldsymbol{r}\cdot(\nabla\boldsymbol{t} \times \boldsymbol{t})\cdot\boldsymbol{r} = 1 \tag{9}$$

with a center at P, which is origin for the vector $\boldsymbol{r}$. Then it is clear from (8) that

The moment of the congruence for any direction at P is equal to the inverse square of the radius of the quadric (9) *in that direction, having the value zero for directions in the asymptotic cone*

$$\boldsymbol{r}\cdot(\nabla\boldsymbol{t} \times \boldsymbol{t})\cdot\boldsymbol{r} = 0. \tag{10}$$

This cone of zero moment was known to Malus, and has been called by Darboux* the *cone of Malus*. It was found by investigating the directions at P which give nearby points, such that the tangent to the curve at one of these points and the tangent at P have a shortest distance apart of the second or higher order. It is the counterpart of the cone (3) of zero tendency. The tangent at P is clearly a generator of the cone of Malus.

Again it follows from (8) and (9) that the sum of the moments of the congruence in three mutually perpendicular directions at P is invariant. The value of this sum is the "scalar"† of $\nabla\boldsymbol{t}\times\boldsymbol{t}$, which is easily shown to have the value $\boldsymbol{t}\cdot\operatorname{rot}\boldsymbol{t}$. Thus:

The sum of the moments of the congruence for any three mutually perpendicular directions at a point is invariant and equal to $\boldsymbol{t}\cdot\operatorname{rot}\boldsymbol{t}$.

This quantity $\boldsymbol{t}\cdot\operatorname{rot}\boldsymbol{t}$ may therefore be called the *total moment* of the congruence at P. It vanishes when the congruence is normal. For, in this case, $\boldsymbol{t}$ is the unit normal to the surface of a singly infinite family, and may therefore be expressed in the form $\psi\nabla\phi$, where ψ and ϕ are point functions. Consequently $\operatorname{rot}\boldsymbol{t} = \nabla\psi\times\nabla\phi$, and

$$\boldsymbol{t}\cdot\operatorname{rot}\boldsymbol{t} = \psi\nabla\phi\cdot(\nabla\psi \times \nabla\phi) = 0.$$

In this case the cone of Malus has an infinite number of sets of three mutually perpendicular generators.‡

* *Leçons sur la Théorie Générale des Surfaces*, vol. 2, pp. 262, 280.

† *Advanced Vector Analysis*, Art. 56.

‡ Darboux, loc. cit.

We shall be concerned largely with the sections of the quadric (9) and the cone of Malus by the normal plane at P. The section of the former is the conic

$$\boldsymbol{r}\cdot\boldsymbol{t} = 0, \quad \boldsymbol{r}\cdot(\nabla\boldsymbol{t}\times\boldsymbol{t})\cdot\boldsymbol{r} = 1 \tag{11}$$

whose asymptotes are the directions of zero moment in the normal plane. For a direction inclined at an angle θ in this plane to a principal axis of the conic (11), the moment is given by

$$M = M_1\cos^2\theta + M_2\sin^2\theta, \tag{12}$$

where M_1 is the moment in the direction of the above axis, and M_2 is that for the perpendicular direction. And clearly $M_1+M_2=\boldsymbol{t}\cdot\mathrm{rot}\ \boldsymbol{t}$, since the direction of $\boldsymbol{t}$ is one of zero moment. It should also be observed that the moment in any direction may be negative. This is the case when the point of the quadric in that direction is imaginary.

We may here pause to draw attention to an interesting result in connection with rectilinear congruences. We have shown elsewhere that, on a ruled surface, the moment M of the family of generators and the Gaussian curvature K at that point are connected by the* relation $M=\pm(-K)^{1/2}$. Consider then any surface of the rectilinear congruence. Let $\boldsymbol{a}$ be the unit vector which is normal to the ray and tangential to the surface. Then the moment in this direction is $\boldsymbol{a}\cdot(\nabla\boldsymbol{t}\times\boldsymbol{t})\cdot\boldsymbol{a}$, and the Gaussian curvature of the ruled surface at P is given by

$$K = -\left[\boldsymbol{a}\cdot(\nabla\boldsymbol{t}\times\boldsymbol{t})\cdot\boldsymbol{a}\right]^2.$$

This vanishes only when $\boldsymbol{a}$ is a direction of zero moment; and if this is so for every point P, the surface is a developable surface of the congruence.

4. Surface of normality. Ultimate surface

If the two directions of zero moment in the normal plane are at right angles, the total moment at that point is zero; and conversely. The locus of points at which the congruence possesses this property may be called the *surface of normality*, for at such points the condition is satisfied that the congruence should be normal. It is the surface of zero total moment, and the points in which it is cut by any curve are the points of zero total moment on that curve. Thus:

---

* See Art. 2 of a paper by the author *On ruled surfaces*, recently communicated to The Mathematical Gazette.

The locus of points at which the two directions of zero moment in the normal plane are at right angles is the surface of normality, or surface of zero total moment. It is given by the equation

$$t \cdot \text{rot}\, t = 0. \tag{13}$$

Corresponding to the limit surface we next consider the locus of points at which the two directions of zero moment in the normal plane are coincident. At such a point the normal plane is tangent to the cone of Malus. The condition for the coincidence of the two directions may be found algebraically by the use of oblique curvilinear coördinates. The method is the same as that adopted in the case of the limit surface in our previous paper* already referred to. As, however, the details of the analysis are rather long, we shall here give only the final result, which may be expressed in invariant form as follows:

The locus of points at which the two directions of zero moment in the normal plane are coincident is the surface given by

$$2\,\text{div}\,(t\,\text{div}\,t - t\cdot\nabla t) - (\text{div}\,t)^2 + (t\cdot\text{rot}\,t)^2 = 0. \tag{14}$$

For convenience we shall refer to this surface as the *ultimate surface*—a name suggested by the term limit surface applied in the corresponding case when the two directions of zero tendency are coincident. We may notice what the theorem becomes in the case of a normal congruence. If J and K are the mean† and Gaussian curvatures of the orthogonal surface, the equation (14) expresses that

$$4K - J^2 = 0,$$

that is to say, *the ultimate surface of a normal congruence is the locus of the umbilical points of the surfaces orthogonal to the congruence.* The interpretation is that at such points the cone of Malus consists of two planes, one of which is perpendicular to t. Any direction in this plane is one of zero moment.

5. Axes of normal sections

We shall next prove the important property that the axes of the conic (4) bisect the angles between those of the conic (11), and vice versa. Take rectangular axes of Cartesian coördinates x, y, z so that the first is in the direction of the tangent t at the point P. Then if i, j, k are unit vectors in the directions of these axes, i is parallel to t, while j and k are perpendicular

* *On congruences of curves*, Art. 4.

† *Differential Geometry*, p. 226 and p. 261.

to t and to each other. The derivatives of t in the directions of the coördinate axes may be expressed in the form

$$a\,j + a'k, \qquad b\,j + b'k, \qquad c\,j + c'k$$

respectively. Then since $r = yj + zk$, the conic (4) is

$$x = 0, \quad by^2 + (b' + c)yz + c'z^2 = 1,$$

and the axes of the conic, being the bisectors of the angles between its asymptotes, are given by

$$\text{(A)} \qquad x = 0, \quad \frac{y^2 - z^2}{b - c'} = \frac{2yz}{b' + c}.$$

Similarly the conic (11) has for its equations

$$x = 0, \quad b'y^2 + (c' - b)yz - cz^2 = 1,$$

and its axes are given by

$$\text{(B)} \qquad x = 0, \quad \frac{y^2 - z^2}{b' + c} = \frac{2yz}{c' - b}.$$

It is clear that the lines (B) are the bisectors of the angles between those given by (A), and vice versa. Hence the theorem:

The axes of the section of either of the quadrics (2) *or* (9) *by the normal plane bisect the angles between the axes of the section of the other.*

This includes as a particular case the theorem for a rectilinear congruence, that the bisectors of the angles between the focal planes are also the bisectors of the angles between the principal planes.* For the cone of Malus is the same at all points of a given ray, consisting of the two focal planes. Thus the planes bisecting the angles between the focal planes contain the axes of the conic (11) for all points of the ray. Similarly it follows from Hamilton's formula† that the principal planes bisect the angles between the asymptotes of the conic (4), and therefore contain the axes of that conic for all points of the ray. From the above theorem it therefore follows that the principal planes are inclined at an angle $\pi/4$ to the bisectors of the angles between the focal planes.

---

* *Differential Geometry*, p. 191.

† *Differential Geometry*, p. 189.

6. Rates of rotation

Consider next the arc-rate at which a direction of zero tendency in the normal plane turns about the tangent as the point moves along the curve. If $\boldsymbol{a}$ is the unit vector in this direction of zero tendency we have identically

$$\boldsymbol{a}\cdot\nabla\boldsymbol{t}\cdot\boldsymbol{a} = 0. \tag{15}$$

Let d/ds denote differentiation along the curve. Then, if $\boldsymbol{b}$ is the unit vector $\boldsymbol{t}\times\boldsymbol{a}$, we may write

$$\frac{d\boldsymbol{a}}{ds} = \omega\boldsymbol{b} + \psi\boldsymbol{t}$$

where ω is the arc-rate of turning about the tangent. Hence on differentiating (15) we have, since $\boldsymbol{t}$ is perpendicular to its derivatives,

$$(\omega\,\boldsymbol{b} + \psi\boldsymbol{t})\cdot\nabla\boldsymbol{t}\cdot\boldsymbol{a} + \boldsymbol{a}\cdot\frac{d}{ds}(\nabla\boldsymbol{t})\cdot\boldsymbol{a} + \omega\boldsymbol{a}\cdot\nabla\boldsymbol{t}\cdot\boldsymbol{b} = 0,$$

which may be written

$$\omega[\boldsymbol{b}\cdot(\nabla\boldsymbol{t}\times\boldsymbol{t})\cdot\boldsymbol{b} - \boldsymbol{a}\cdot(\nabla\boldsymbol{t}\times\boldsymbol{t})\cdot\boldsymbol{a}] = \boldsymbol{a}\cdot\frac{d}{ds}(\nabla\boldsymbol{t})\cdot\boldsymbol{a} + \psi\boldsymbol{t}\cdot\nabla\boldsymbol{t}\cdot\boldsymbol{a}. \tag{16}$$

At the limit surface of the congruence the coefficient of ω is zero. For $\boldsymbol{a}$ and $\boldsymbol{b}$ have then the directions of the axes of the conic (4), and are therefore inclined at equal angles $\pi/4$ to each of the axes of the conic (11). It follows from (12) that the moments in these directions are equal, showing that the coefficient of ω vanishes at the limit surface. To interpret this, we observe that, just as ω is the arc-rate of rotation of the direction of zero tendency in the normal plane as the point P moves along the curve, so its reciprocal is the rate at which the point P moves along the curve for rotation of the normal direction of zero tendency. Thus, since the coefficient of ω in (16) vanishes at a limit point, while in general the second member of (16) does not, we have the following theorem:

At the limit points of a curve the feet of the normals giving the directions of zero tendency in the normal plane are stationary for variation of these directions.

This is substantially the theorem found in our earlier paper by a different method, and stated in terms of common normals to the curve and nearby curves.*

Similarly we may examine the rate of rotation of a direction of zero

* *On congruences of curves*, Art. 4.

moment in the normal plane. If $\boldsymbol{a}$ is now the unit vector in such a direction we have identically

$$\boldsymbol{a}\cdot(\nabla\boldsymbol{t}\times\boldsymbol{t})\cdot\boldsymbol{a}=0.$$

Hence, with the same notation as before, we have on differentiation

$$(\omega\boldsymbol{b}+\psi\boldsymbol{t})\cdot(\nabla\boldsymbol{t}\times\boldsymbol{t})\cdot\boldsymbol{a}+\boldsymbol{a}\cdot\frac{d}{ds}(\nabla\boldsymbol{t}\times\boldsymbol{t})\cdot\boldsymbol{a}+\omega\boldsymbol{a}\cdot(\nabla\boldsymbol{t}\times\boldsymbol{t})\cdot\boldsymbol{b}=0,$$

which may be expressed as

$$\omega[\boldsymbol{a}\cdot\nabla\boldsymbol{t}\cdot\boldsymbol{a}-\boldsymbol{b}\cdot\nabla\boldsymbol{t}\cdot\boldsymbol{b}]=\boldsymbol{a}\cdot\frac{d}{ds}(\nabla\boldsymbol{t}\times\boldsymbol{t})\cdot\boldsymbol{a}+\psi\boldsymbol{t}\cdot(\nabla\boldsymbol{t}\times\boldsymbol{t})\cdot\boldsymbol{a}. \tag{17}$$

At the ultimate surface the coefficient of ω is zero. For $\boldsymbol{a}$ and $\boldsymbol{b}$ have then the directions of the axes of the conic (11), and are therefore inclined at equal angles $\pi/4$ to those of the conic (4). It follows that the tendencies in these directions are equal, so that the coefficient of ω in (17) vanishes. In general the second member of (17) is not zero, and we have the following theorem:

At the ultimate points of a curve the feet of the normals giving the directions of zero moment in the normal plane are stationary for variation of these directions.

7. Normal congruence

It is known that, for a normal congruence of straight lines, the foci coincide with the limits.* In order to extend the theorem to a normal congruence of curves, we shall make use of the system of oblique curvilinear coördinates adopted in our earlier paper. Let s be the distance measured along a curve from a given surface, called the *director surface*, while u, v are current parameters for a point on that surface. Any curve is determined by the values of u, v for the point at which it crosses the director surface. Let $\boldsymbol{r}$ be the position vector of a point in space, and suffixes 1, 2, 3 denote differentiation with respect to u, v, s respectively. Then the unit tangent $\boldsymbol{t}$ is $\boldsymbol{r}_3$.

Consider a point on the curve (u, v) whose distance from a point on a nearby curve $(u+du, v+dv)$ is of the second or higher order. Then, if $\boldsymbol{r}(u, v, s)$ and $\boldsymbol{r}(u+du, v+dv, s+ds)$ are these points on the two curves, we have to the first order

$$\boldsymbol{r}(u,v,s)=\boldsymbol{r}(u+du,v+dv,s+ds)$$

so that

$$\boldsymbol{r}_1\,du+\boldsymbol{r}_2\,dv+\boldsymbol{r}_3\,ds=0, \tag{18}$$

* *Differential Geometry*, Art. 100.

showing that r_1, r_2, r_3 are coplanar, and leading to the equation of the focal surface $[r_1, r_2, r_3]=0$.

Since t is perpendicular to its derivatives, it follows on forming the scalar product of the members of (18) with t_1, t_2, t_3 in turn that the quotient du/dv has the value

$$\frac{du}{dv} = -\frac{r_2\cdot t_1}{r_1\cdot t_1} = -\frac{r_2\cdot t_2}{r_1\cdot t_2} = -\frac{r_2\cdot t_3}{r_1\cdot t_3},$$

and consequently, for points on the focal surface,

$$\begin{aligned}(r_2\cdot t_1)(r_1\cdot t_3) &= (r_1\cdot t_1)(r_2\cdot t_3),\\ (r_2\cdot t_2)(r_1\cdot t_3) &= (r_1\cdot t_2)(r_2\cdot t_3).\end{aligned} \tag{19}$$

In terms of the magnitudes

$$\begin{aligned}a &= r_1^2, \qquad b = r_2^2,\\ f &= r_2\cdot r_3, \quad g = r_3\cdot r_1, \quad h = r_1\cdot r_2\end{aligned}$$

the first of (19) may be expressed by

$$(f_1 - g_2 + h_3)g_3 = a_3f_3. \tag{20}$$

Now for a normal congruence $t\cdot \operatorname{rot} t$ vanishes identically. This may be expressed* by

$$f_1 - g_2 = gf_3 - fg_3.$$

Substituting this value in (20) we have

$$(gf_3 - fg_3 + h_3)g_3 = a_3f_3,$$

or

$$f_3B_3 + g_3H_3 = 0, \tag{21}$$

A, B, H being the cofactors of a, b, h respectively in the determinant

$$\begin{vmatrix} a & h & g\\ h & b & f\\ g & f & 1\end{vmatrix}.$$

Similarly from the second of (19) we find

$$g_3A_3 + f_3H_3 = 0. \tag{22}$$

From this equation and (21) it then follows that

$$A_3B_3 = H_3^2, \tag{23}$$

* *On congruences of curves*, Art. 8.

which is the equation of the limit surface* of the congruence. Thus at the focal points of a curve the equation of the limit surface is satisfied, and we have the following theorem:

For a normal congruence of curves the foci lie on the limit surface.

8. Where the cones are pairs of planes

We have seen that the cone of Malus consists of a pair of planes at all points of a rectilinear congruence, and also at points on the ultimate surface of a normal curvilinear congruence. Let us consider at what points either of the two cones thus becomes a pair of planes. Since the tangent at the point is a generator of each cone, it is clear that when the cone consists of a pair of planes, one of these must pass through the tangent.

First consider the cone of zero tendency, and suppose that this is a pair of planes. Let $\boldsymbol{a}$ be the unit vector parallel to the intersection of the plane through the tangent with the normal plane at P. Then any direction at P in the plane of $\boldsymbol{t}$ and $\boldsymbol{a}$ is one of zero tendency; so that, for all values of ϕ,

$$(\boldsymbol{t}\cos\phi + \boldsymbol{a}\sin\phi)\cdot\nabla\boldsymbol{t}\cdot(\boldsymbol{t}\cos\phi + \boldsymbol{a}\sin\phi) = 0.$$

Since $\boldsymbol{t}$ is a unit vector it is perpendicular to its derivatives. Also $\boldsymbol{a}$ is by hypothesis a direction of zero tendency, so that the above condition requires that

$$\boldsymbol{t}\cdot\nabla\boldsymbol{t}\cdot\boldsymbol{a} = 0, \tag{24}$$

which may be expressed as

$$\kappa\boldsymbol{n}\cdot\boldsymbol{a} = 0,$$

where κ is the curvature of the curve, and $\boldsymbol{n}$ the unit principal normal. This equation is satisfied if κ is zero, or if $\boldsymbol{a}$ is the unit binormal to the curve. Hence:

At points where the curvature of a curve is zero, or where the binormal is a direction of zero tendency, the cone of zero tendency consists of a pair of planes.

The former condition is satisfied at all points for a rectilinear congruence; and in this case the cone of zero tendency at any point of a ray is a pair of planes equally inclined to each principal plane.

Next consider the cone of Malus. In order that this may consist of a pair of planes, we must have, for all values of ϕ,

$$(\boldsymbol{t}\cos\phi + \boldsymbol{a}\sin\phi)\cdot(\nabla\boldsymbol{t}\times\boldsymbol{t})\cdot(\boldsymbol{t}\cos\phi + \boldsymbol{a}\sin\phi) = 0,$$

* *On congruences of curves*, Art. 4.

$\boldsymbol{a}$ being the unit vector in the normal plane and in the plane of zero moment through the tangent. This requires that

$$\boldsymbol{t}\cdot(\nabla\boldsymbol{t}\times\boldsymbol{t})\cdot\boldsymbol{a}=0, \tag{25}$$

which may be expressed as

$$\kappa\boldsymbol{n}\times\boldsymbol{t}\cdot\boldsymbol{a}=0.$$

Now $\boldsymbol{t}\times\boldsymbol{n}$ is the unit binormal to the curve. Hence, in order that (25) may hold, either κ must vanish, or $\boldsymbol{a}$ must be parallel to $\boldsymbol{n}$. Thus:

At points where the curvature of a curve is zero, or where the principal normal is a direction of zero moment, the cone of Malus consists of a pair of planes.

The former condition is satisfied at all points for a rectilinear congruence; and in this case the cone of Malus consists of the two focal planes, which are the same for all points of a given ray.

9. Isotropic congruence

The conception of an isotropic congruence* of straight lines may be generalised so as to apply to curvilinear congruences. An isotropic rectilinear congruence is one whose limit surface coincides with its surface of striction. Now at points on the former surface the normal plane is tangent to the cone of zero tendency, while at points on the latter the sum of the tendencies in two perpendicular directions in the normal plane is zero. In order that both properties may be possessed simultaneously, the cone of zero tendency must consist of two planes, one of which is the normal plane. Conversely, points at which the cone (3) behaves in this manner must lie on both the limit surface and the surface of striction. We shall therefore define an *isotropic curvilinear congruence* as one for which the normal plane is part of the cone of zero tendency at all points of the surface of striction. This surface may then be described as a limit-striction surface. At points on it the equation (7) holds simultaneously with div $\boldsymbol{t}=0$.

Let $\boldsymbol{a}$ and $\boldsymbol{b}$ be a pair of perpendicular unit vectors in the normal plane. Then, at points where the above property holds, the direction of the unit vector $\boldsymbol{a}\cos\phi+\boldsymbol{b}\sin\phi$ must be one of zero tendency for all values of ϕ. Hence the equation

$$(\boldsymbol{a}\cos\phi+\boldsymbol{b}\sin\phi)\cdot\nabla\boldsymbol{t}\cdot(\boldsymbol{a}\cos\phi+\boldsymbol{b}\sin\phi)=0$$

is equivalent to the three relations

* *Differential Geometry*, Art. 102.

$$\boldsymbol{a}\cdot\nabla\boldsymbol{t}\cdot\boldsymbol{a}=0, \qquad \boldsymbol{b}\cdot\nabla\boldsymbol{t}\cdot\boldsymbol{b}=0,$$

and

$$\boldsymbol{a}\cdot(\nabla\boldsymbol{t}\times\boldsymbol{t})\cdot\boldsymbol{a}-\boldsymbol{b}\cdot(\nabla\boldsymbol{t}\times\boldsymbol{t})\cdot\boldsymbol{b}=0.$$

From the last of these it follows that the moments in the directions of $\boldsymbol{a}$ and $\boldsymbol{b}$ are equal. Since this must hold for all pairs of perpendicular directions in the normal plane, the conic (11) must be a circle, real or imaginary, and the moment must be the same for all directions in that plane, having the value $\frac{1}{2}\boldsymbol{t}\cdot\text{rot }\boldsymbol{t}$. Thus at a point P on this limit-striction surface, the moment of the family of curves on any surface of the congruence has this same value.* The directions of zero moment in the normal plane are therefore minimal. Also in the particular case of a rectilinear congruence, the Gaussian curvature K has the same value at P for all the ruled surfaces of the congruence, being equal† to $-\frac{1}{4}(\boldsymbol{t}\cdot\text{rot }\boldsymbol{t})^2$, and the parameter of distribution of the ray is the same for all such surfaces.‡

For an isotropic congruence of curves the intersection of the limit-striction surface with any surface of the congruence is the line of striction§ of the family of curves on that surface; for it is the locus of points at which the tendency is zero for the direction perpendicular to the curve.

Similarly we might imagine a curvilinear congruence with the property that, at all points of the surface of normality, the cone of Malus consists of a pair of planes, one of which is the normal plane to the curve. Such points must lie also on the ultimate surface; and the coalescence of these two surfaces gives what may be called the ultimate-normality surface. As before, let $\boldsymbol{a}$ and $\boldsymbol{b}$ be perpendicular unit vectors in the normal plane. Then if this plane is part of the cone of zero moment, we must have

$$(\boldsymbol{a}\cos\phi+\boldsymbol{b}\sin\phi)\cdot(\nabla\boldsymbol{t}\times\boldsymbol{t})\cdot(\boldsymbol{a}\cos\phi+\boldsymbol{b}\sin\phi)=0$$

for all values of ϕ. This is equivalent to

$$\boldsymbol{a}\cdot(\nabla\boldsymbol{t}\times\boldsymbol{t})\cdot\boldsymbol{a}=0, \qquad \boldsymbol{b}\cdot(\nabla\boldsymbol{t}\times\boldsymbol{t})\cdot\boldsymbol{b}=0,$$

and

$$\boldsymbol{b}\cdot\nabla\boldsymbol{t}\cdot\boldsymbol{b}-\boldsymbol{a}\cdot\nabla\boldsymbol{t}\cdot\boldsymbol{a}=0.$$

From the last of these it follows that the tendencies in the directions of $\boldsymbol{a}$ and $\boldsymbol{b}$ are equal; and since this is true for all pairs of perpendicular vectors, the conic (4) must be a circle, real or imaginary. Thus the tendency has the same value, $\frac{1}{2}$ div $\boldsymbol{t}$, for all directions in the normal plane, being also equal

---

* *Differential Geometry*, Art. 130.
† *On ruled surfaces*, Art. 2.
‡ *Differential Geometry*, p. 198.
§ Ibid., p. 249.

to the divergence* at that point of the family of curves on any surface of the congruence through the point.

For a congruence of this nature the directions of zero tendency in the normal plane are minimal at points on the ultimate-normality surface. Also the intersection of this surface with any surface of the congruence is a line of zero moment for the family of curves on the latter,† for the moment of this family is zero for the direction in the surface perpendicular to the curve.

10. Three orthogonal congruences

Finally let us consider briefly some properties of three curvilinear congruences cutting one another orthogonally at all points. Let $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$ be the unit tangents to the curves at any point, so that

$$\boldsymbol{a} = \boldsymbol{b} \times \boldsymbol{c}, \qquad \boldsymbol{b} = \boldsymbol{c} \times \boldsymbol{a}, \qquad \boldsymbol{c} = \boldsymbol{a} \times \boldsymbol{b}.$$

The total moments A, B, C of the three congruences have the values

$$A = \boldsymbol{a} \cdot \operatorname{rot} \boldsymbol{a}, \qquad B = \boldsymbol{b} \cdot \operatorname{rot} \boldsymbol{b}, \qquad C = \boldsymbol{c} \cdot \operatorname{rot} \boldsymbol{c}.$$

From the last equation we have

$$C = \boldsymbol{c} \cdot \operatorname{rot} (\boldsymbol{a} \times \boldsymbol{b}) = \boldsymbol{c} \cdot (\boldsymbol{b} \cdot \nabla \boldsymbol{a} - \boldsymbol{a} \cdot \nabla \boldsymbol{b} + \boldsymbol{a} \operatorname{div} \boldsymbol{b} - \boldsymbol{b} \operatorname{div} \boldsymbol{a})$$

or

$$C = \boldsymbol{b} \cdot (\nabla \boldsymbol{a} \times \boldsymbol{a}) \cdot \boldsymbol{b} + \boldsymbol{a} \cdot (\nabla \boldsymbol{b} \times \boldsymbol{b}) \cdot \boldsymbol{a}. \tag{26}$$

Also since $\boldsymbol{a}$ is perpendicular to $\boldsymbol{b}$ it follows that

$$0 = \nabla(\boldsymbol{a} \cdot \boldsymbol{b}) = \boldsymbol{a} \cdot \nabla \boldsymbol{b} + \boldsymbol{b} \cdot \nabla \boldsymbol{a} + \boldsymbol{a} \times \operatorname{rot} \boldsymbol{b} + \boldsymbol{b} \times \operatorname{rot} \boldsymbol{a},$$

and consequently, on forming the scalar product with $\boldsymbol{b} \times \boldsymbol{a}$, we have

$$0 = \boldsymbol{a} \cdot (\nabla \boldsymbol{b} \times \boldsymbol{b}) \cdot \boldsymbol{a} - \boldsymbol{b} \cdot (\nabla \boldsymbol{a} \times \boldsymbol{a}) \cdot \boldsymbol{b} - B + A. \tag{27}$$

From this equation and (26) it follows that

$$\begin{aligned} \boldsymbol{a} \cdot (\nabla \boldsymbol{b} \times \boldsymbol{b}) \cdot \boldsymbol{a} &= \tfrac{1}{2}(B + C - A), \\ \boldsymbol{b} \cdot (\nabla \boldsymbol{a} \times \boldsymbol{a}) \cdot \boldsymbol{b} &= \tfrac{1}{2}(A - B + C). \end{aligned} \tag{28}$$

Similarly we may show that

$$\begin{aligned} \boldsymbol{a} \cdot (\nabla \boldsymbol{c} \times \boldsymbol{c}) \cdot \boldsymbol{a} &= \tfrac{1}{2}(B + C - A), \\ \boldsymbol{c} \cdot (\nabla \boldsymbol{a} \times \boldsymbol{a}) \cdot \boldsymbol{c} &= \tfrac{1}{2}(A + B - C). \end{aligned} \tag{29}$$

From (28) and (29) it follows that the moments of the second and third congruences in the direction of the first have the same value. If we write

$$2S = A + B + C$$

we may state the above results as follows:

* *Differential Geometry*, p. 259 and Art. 116.

† Ibid., p. 258.

The moments of the second and third congruences in the direction of the first have each the value $S-A$; those of the first and third in the direction of the second have the value $S-B$, and those of the first and second in the direction of the third the value $S-C$.

Thus the moments of the three orthogonal congruences are not independent, but are connected by the relations expressed in this theorem.

An important particular case is that in which two of the congruences are normal, say the first and second. Then A and B both vanish; and the curves of the third congruence are the lines of intersection of two orthogonal families of surfaces, whose unit normals are $\boldsymbol{a}$ and $\boldsymbol{b}$ respectively. The moment of the third congruence in either of the directions $\boldsymbol{a}$ or $\boldsymbol{b}$ is equal to $\frac{1}{2}C$; that is to say:

If the curves of a congruence are the lines of intersection of two orthogonal families of surfaces, the moments of the congruence in the directions of the normals to these surfaces are equal to each other and to half the total moment of the congruence.

The theorem just enunciated has several known theorems as immediate corollaries. The moment of the congruence in the direction of $\boldsymbol{b}$ is the moment of the family of curves on the surface whose normal is $\boldsymbol{a}$; and this vanishes identically only when these curves are lines of curvature on the surface.* Similarly the moment of the congruence in the direction of $\boldsymbol{a}$ is that of the family of curves on the surface whose normal is $\boldsymbol{b}$. Hence the theorem that if the curves of intersection are lines of curvature on one of the orthogonal families of surfaces, they are lines of curvature also on the other.† In that case C is zero, and the curves of intersection constitute a normal congruence. Thus we have Darboux's theorem that there is a third family of surfaces orthogonal to the first two.‡ Conversely, if C is zero, the moment of the family of curves on each of the orthogonal surfaces is zero, and the curves are therefore lines of curvature. Hence we have Dupin's theorem that the curves of intersection of the surfaces of a triply orthogonal system are lines of curvature on these surfaces.§

---

* *Differential Geometry*, Art. 130.
† Ibid., Art. 112.
‡ Darboux, loc. cit., p. 276.
§ *Differential Geometry*, Art. 109.

CANTERBURY COLLEGE,
CHRISTCHURCH, NEW ZEALAND

THE MULTINOMIAL SOLID AND THE CHI TEST*

BY

B. H. CAMP

1. INTRODUCTION

Let p_i be the relative frequency of $x_i (i=1, 2, \cdots, m)$ in a frequency distribution. If the number of elements in the distribution be infinite and a sample of size n be drawn, the probability of obtaining exactly s_1 elements of size x_1, and at the same time s_2 of size x_2, etc., is known to be

$$(1) \qquad P = \frac{n!}{s_1! \cdots s_m!} p_1^{s_1} \cdots p_m^{s_m}, \qquad p_1 + p_2 + \cdots + p_m = 1,$$

where the s's are integers whose sum is n. P is a function of $m-1$ independent variables, and is called the multinomial solid. When $m=2$, it degenerates into the general term of the point binomial, $(p_1+p_2)^n$. Although much has been known about this solid, its usefulness has not been sufficiently recognized. In this paper I shall point out certain of its properties and two applications, one to a well known theorem in sampling, and one to the χ-test. The proof of the χ-test here given may be thought of as supplementing Pearson's original proof† in that it replaces an assumed approximation by a demonstrated approximation. It is not related to Pearson's second proof‡ which is a variant of Soper's.§

2. GENERAL PROPERTIES¶

(a) P may be written in the form $P=P_1 \cdots P_m=P_1 \cdots P_{m-1}$; where P_1 is a function of s_1 only, and is a point binomial; P_2 is a function of s_1, s_2 only, and is a point binomial if s_1 be fixed; etc.; P_i is a function of $s_1 \cdots s_i$ only, and is a point binomial if $s_1 \cdots s_{i-1}$ be fixed; $P_m=1$; thus:

---

* Presented to the Society, December 29, 1926; received by the editors in June, 1928.

† K. Pearson, *On deviations from the probable in random sampling*, Philosophical Magazine, (5), vol. 50 (1900), pp. 157–175.

‡ K. Pearson, *On the general theory of multiple contingency*, Biometrika, vol. 11 (1916), pp. 145–158.

§ Soper, *Frequency Arrays*, Cambridge University Press, 1922, p. 15, footnote.

This proof makes other approximations which do not conveniently lend themselves to measurement. I shall be particularly interested here in the degree of approximation attained in the application of the chi test to ordinary problems.

¶ The properties grouped in this section are not essentially new.

$$P = \left[\frac{n!}{s_1!(n-s_1)!}p_1^{s_1}(1-p_1)^{n-s_1}\right]\left[\frac{(n-s_1)!}{s_2!(n-s_1-s_2)!}\left(\frac{p_2}{1-p_1}\right)^{s_2}\cdot\left(\frac{1-p_1-p_2}{1-p_1}\right)^{n-s_1-s_2}\right]$$

$$\left[\frac{(n-s_1-s_2)!}{s_3!(n-s_1-s_2-s_3)!}\left(\frac{p_3}{1-p_1-p_2}\right)^{s_3}\left(\frac{1-p_1-p_2-p_3}{1-p_1-p_2}\right)^{n-s_1-s_2-s_3}\right]\cdots$$

$$\cdot\left[\frac{s_m!}{s_m!(s_m-s_m)!}\left(\frac{p_m}{1-p_1-\cdots-p_{m-1}}\right)^{s_m}=1\right].$$

(b) The form (a) shows that equation (1) might have been derived by the following reasoning: Find the probability of exactly s_1 successes with p_1 as probability and n as the number of drawings; multiply by the probability of exactly s_2 successes with $p_2/(1-p_1)$ as probability and $(n-s_1)$ as the number of drawings; multiply, etc.

Definitions. (α) Any positive integrable function of $(m-1)$ independent variables may be considered an $(m-1)$-*way frequency solid* (or surface or curve), and may be represented in m-dimensional space.

(β) If, in an $(m-1)$-way solid, r variables be fixed, there is left a function of $(m-1-r)$ variables, therefore an $(m-1-r)$-way solid; this is to be called an $(m-1-r)$-*way array*.

Example. If $m=4$, the $(m-1)$-way solid is a 3-way correlation solid, say a function of x, y, and z. If one fixes x and y, thus taking $r=2$, the one-way arrays resulting are straight lines or columns in the z-direction. If, further, one added up the frequencies in each of these columns, one would arrive at a set of totals which might be thought of as spread over the xy-plane, thus comprising a two-way correlation table.

(γ) If one adds the frequencies in each of the $(m-1-r)$-way arrays, the result is a function of r variables; this is an r-*way total*.

(c) **Totals.** If, in P above, the variable s_1 be fixed, and if the frequencies of each of the $(m-2)$-way arrays thus determined be added, the one-way total finally determined is P_1. If both the variables s_1 and s_2 be fixed, and if the frequencies in each of the $(m-3)$-way arrays thus determined be added, the two-way total determined is P_1P_2. If $s_1, s_2, \cdots, s_{m-2}$ be fixed, and the totals of the one-way arrays be found, the $(m-2)$-way total that results is $P_1P_2\cdots P_{m-2}$.

These facts follow easily from the form (a). E.g., fixing s_1,

$$\sum_{s_2}\sum_{s_3}\cdots\sum_{s_{m-1}}P_1\cdots P_{m-1}=P_1\left(\sum_{s_2}P_2\right)\cdots\left(\sum_{s_{m-1}}P_{m-1}\right)=P_1.$$

(d) **Means of totals.** The mean of the one-way total of P in the s_1 direction is p_1n; the mean of the two-way total in s_1s_2-space is (p_1n, p_2n); the general mean of the whole solid is $(p_1n, p_2n, \cdots, p_{m-1}n)$.

These facts follow also from the form (a) and the well known theorems that the coördinates of the general mean of a correlation solid are the means of its one-way totals.

3. Regression surfaces

(a) The mean of the one-way array in the s_{m-1} direction made by fixing $s_1, s_2, \cdots, s_{m-2}$ is

$$\frac{p_{m-1}}{1 - p_1 - \cdots - p_{m-2}}(n - s_1 - s_2 - \cdots - s_{m-2}).$$

This may be thought of as a linear function of the $m-2$ variables, $s_1, s_2, \cdots, s_{m-2}$, and thus yields the equation of the regression surface, s_{m-1} on $(s_{m-2}, \cdots, s_1)$.

Use the form of §2(a). The one-way array is $P_1P_2 \cdots P_{m-1}$, and its total is, by §2(c), $P_1P_2 \cdots P_{m-2}$. So its mean is

$$\frac{1}{P_1 \cdots P_{m-2}} \sum_{s_{m-1}} s_{m-1}P_1 \cdots P_{m-1} = \sum_{s_{m-1}} s_{m-1}P_{m-1}.$$

The last named sum is the mean number of successes in a point binomial distribution in which $p_{m-1}/(1-p_1- \cdots -p_{m-2})$ is the probability of a success, and $(n-s_1- \cdots -s_{m-2})$ is the number of drawings. This is well known to have the value given above.

(b) Every partial regression is also a linear function. (For one needs only to fix certain of the variables of the regression surface in (a) to obtain the partial regression surface.) It follows from (a) and (b) that, for $m=4$, this solid is an example of the all linear case I have considered elsewhere.*

Application. Let $m=3$. The mean of the one-way array made by fixing s_1 is the following linear function of s_1: $p_2(n-s_1)/(1-p_1)$. This fact, now proved to be true, was chosen by Pearson as a reasonable assumption in his proof of a fundamental theorem in sampling,† viz.: that the coefficient of correlation between two frequencies with which two fixed values of a variable will occur is $-(p_1p_2/(q_1q_2))^{1/2}$.

---

* Camp, *Mutually consistent multiple regression surfaces*, Biometrika, vol. 17 (1925), p. 450.

† K. Pearson, *On the probable errors of frequency constants*, Biometrika, vol. 2 (1902–3), p. 274. It has been copied, as an unproved assumption, in several texts.

4. The Laplace approximation

It has been generally assumed, specifically by Pearson in deriving the χ-test, that the multinomial solid just considered could be approximated when n is large, by the normal solid. I wish to examine the measure of this approximation, and to do so I shall also derive the normal solid from the multinomial solid in a manner similar to that used by Laplace for the one-dimensional case.

In (1) of §1, put $s_i = np_i + k_i$, $i = 1, 2, \cdots, m$. Then P becomes

$$(2) \qquad P_{k_1 \cdots k_m} = \frac{n!}{(np_1 + k_1)! \cdots (np_m + k_m)!} p_1^{np_1 + k_1} \cdots p_m^{np_m + k_m},$$

where

$$(3) \quad p_1 + \cdots + p_m = 1\,; \quad k_1 + \cdots + k_m = 0, \text{ since } s_1 + \cdots + s_m = n\,; \text{ or}$$

$$(4) \qquad P = n!(r_1 \cdots r_m), \quad \text{where} \quad r_i = p_i^{np_i + k_i}/(np_i + k_i)!.$$

Let

$$(5) \qquad \epsilon = 1/(np), \quad \omega = k\epsilon = k/(np), \quad \alpha = k\omega = k^2/(np).$$

Use Stirling's approximation, and the symbol $\cong$ to mean "is approximately equal to," and drop the subscript from r temporarily:

$$\begin{aligned} \log r &\cong \log p^{np+k} - (np + k + \tfrac{1}{2}) \log (np + k) + (np + k) - \log (2\pi)^{1/2} \\ &= \log (p^{np(1+\omega)}/(2\pi)^{1/2}) - [np(1+\omega) + \tfrac{1}{2}][\log np + \omega - \omega^2/2 - \omega^3/3 - \cdots] \\ &\qquad + np(1 + \omega) \\ &= \log (p^{np(1+\omega)}/(2\pi)^{1/2}) - \log (np)^{np(1+\omega)+1/2} \\ &\qquad - [np(1 + \omega) + \tfrac{1}{2}][\omega - \omega^2/2 + \omega^3/3 - \cdots] + np(1 + \omega) \\ &= A + \log p^{np(1+\omega)} - \log (2n\pi p)^{1/2} (np)^{np(1+\omega)}, \end{aligned}$$

where

$$A = np(1 + \omega)(1 - \omega + \omega^2/2 - \omega^3/2 + \cdots) + \tfrac{1}{2}(-\omega + \omega^2/2 - \omega^3/3 + \cdots).$$

So

$$(6) \qquad r \cong e^A [(2n\pi p)^{1/2} n^{np(1+\omega)}]^{-1},$$

and the relative error thus introduced into r is known* to be $1 - e^{-t/12}$, where $0 < t < 1/(np(1+\omega))$. It is now necessary to impose certain restrictions on ϵ, ω, α, m. These will seem arbitrary to the reader, and indeed they are.

---

* The usual statement is $x! = (2\pi x)^{1/2} x^x e^{-x} e^{\theta/(12x)}$, where $0 < \theta < 1$. Hence, if y is the approximate value, $y = x! e^{-\theta/(12x)}$. The relative error in $x!$ is $(-y + x!)/x! = -y/x! + 1 = -e^{-\theta/(12x)} + 1$, where $0 < \theta/x < 1/x$.

They have been chosen after considerable experimentation so as to give results as favorable as possible to the general purposes of the χ-test. Doubtless more useful conditions might be found. Certainly it would be easy to discover better conditions if one restricted his view to special cases. For example, the point binomial case ($m=2$) is a very special one. These conditions are not chosen so as to measure well the closeness of the approximation in this case. Let

$$(7)\qquad \text{(a) } np - |k| \geqq 3, \qquad \text{(b) } |k|/(np) = |\omega| < 0.2,$$
$$\text{(c) } k^2/(np) = \alpha < 6, \qquad \text{(d) } m \leqq 20.$$

It follows that $np(1+\omega)=np+k\geqq 3$, and so $t<1/(np(1+\omega))<1/3$; and therefore the maximum relative error in r, as given by (6), is

$$(8)\qquad \delta_1 = 1 - e^{-t/12} \leqq 0.09202/(np+k).$$

Now let $A=np(1-\omega^2/2)+B$, where

$$(9)\qquad B = \omega\alpha\left(\frac{1}{3\cdot 2} - \frac{\omega}{4\cdot 3} + \frac{\omega^2}{5\cdot 4} - \cdots\right) + \frac{\omega}{2}\left(-1 + \frac{\omega}{2} - \frac{\omega^2}{3} + \cdots\right)$$
$$= (\omega/6)(\alpha-3)(1-\omega/2) + \omega\alpha(\omega^2/20 - \omega^3/30 + \omega^4/42 - \cdots)$$
$$+ (\omega^3/2)(-1/3 + \omega/4 - \omega^2/5 + \cdots).$$

So

$$(10)\qquad e^A = e^{np(1-\omega^2/2)}(e^B \cong 1+B),$$

and the error thus made in e^B is $(B^2/2)(1+B/3+B^2/12+\cdots)$.

We can find an upper limit to this by the use of (9) and (7) (b) and (c):

$$|B| < \left|\frac{\omega}{6}(\alpha-3)\left(1-\frac{\omega}{2}\right)\right| + \left|\omega^3\left\{\left(\frac{\alpha}{20}-\frac{1}{6}\right)\right.\right.$$
$$\left.\left. - \omega\left(\frac{\alpha}{30}-\frac{1}{8}\right) + \omega^2\left(\frac{\alpha}{42}-\frac{1}{10}\right) - \omega^3\left(\frac{\alpha}{56}-\frac{1}{12}\right) + \cdots\right\}\right|$$
$$< |(\omega/6)(\alpha-3)(1-\omega/2)|$$
$$+ |\omega^3|\{1/6 + |\omega/8| + \omega^2/10 + |\omega^3/12| + \cdots\}$$
$$< (1/30)(3)(1.1) + 0.0014 < 0.1114.$$

$$(11)\qquad \text{Let } a = \omega(\alpha-3)/6. \quad \text{By (7)}, \quad |a| < 0.1.$$

$$(12)\quad B^2(1+|B|/3)/2 < \{a^2(1-\omega/2)^2$$
$$+ 0.000,308 + 0.000,001,96\}(1 + 0.0372 + 0.00104 + 0.0001 + \cdots)/2$$
$$< \{a^2(1-\omega/2)^2 + 0.000,155\}(0.520) \cong 0.52a^2(1-\omega/2)^2,$$

which we shall take as the maximum absolute error made in e^B by the approximation of (10). We shall now increase this error slightly by cutting off from (10) a part of B, writing $B \cong a$. The additional absolute error thus made is, by (11) and (9), at most equal to

$$\begin{aligned}&|\,\omega\alpha(-\omega/12+\omega^2/20-\omega^3/30+\omega^4/42-\cdots)\\&\qquad+(\omega/2)(\omega/2-\omega^2/3+\omega^3/4-\omega^4/5+\omega^5/6-\cdots)\,|\\&=\left|\frac{\omega^2}{4}\left(1-\frac{\alpha}{3}\right)+\frac{\omega^3}{2}\left(\frac{\alpha}{10}-\frac{1}{3}\right)+\frac{\omega^4}{2}\left(-\frac{\alpha}{15}+\frac{1}{4}\right)+\frac{\omega^5}{2}\left(\frac{\alpha}{21}-\frac{1}{5}\right)+\cdots\right|\\&<\omega^2/4+|\,\omega\,|^3/6+\omega^4/8+|\,\omega\,|^5/10+\cdots.\end{aligned}$$

From this inequality and (12) it follows that the maximum error made in e^B by writing

$$e^B \cong 1+a \tag{13}$$

is

$$0.52a^2(1-\omega/2)^2+\omega^2/4+|\,\omega\,|^3/6+\omega^4/8+\cdots. \tag{14}$$

We need, however, the maximum *relative* error in e^B. To get this we divide (14) by the minimum value of e^B, which is the series $1-0.1114+(0.1114)^2/2-(0.1114)^3/6+\cdots=0.8945$. Hence the maximum relative error in e^B as given by (13) is

$$\begin{aligned}\delta_2&=0.582a^2(1-\omega/2)^2+0.559(\omega^2/2+|\,\omega\,|^3/3+\omega^4/4+\cdots)\\&<0.582(\omega^2/4)(1+|\,\omega\,|+\omega^2/4)+0.559(\omega^2/2+|\,\omega\,|^3/3+\omega^4/4+\cdots)\\&=0.425\omega^2+0.332|\,\omega\,|^3+0.176\omega^4+\cdots<0.02.\end{aligned} \tag{15}$$

By (6), (10), and (13) we have now made altogether the following approximation to r:

$$r \cong \frac{(1+a)e^{np(1-\omega^2/2)}}{(2\pi pn)^{1/2}n^{np(1+\omega)}}. \tag{16}$$

By (8) and (15) the relative error in r thus made is at most

$$\begin{aligned}\rho&=\delta_1+\delta_2+\delta_1\delta_2\\&<0.094/(np-|\,k\,|)+0.425\omega^2+0.332|\,\omega\,|^3+0.176\omega^4+\cdots,\end{aligned} \tag{17}$$

and we shall hereafter omit the terms not printed in this series. The following known theorem has just been used and will be used again shortly: Let δ_i be the relative error in x_i, $i=1,\cdots,m$. The relative error in the product $x_1\cdots x_m$ is

$$\sum\delta_i-\sum\sum\delta_i\delta_j+\sum\sum\sum\delta_i\delta_j\delta_k-\cdots, \qquad \text{where } i\neq j\neq k, \text{ etc.}$$

It now follows that the maximum absolute relative error in the product $r_1 \cdots r_m$ of equation (4) is

$$\sum_1^m \rho_i + m(m-1)(0.0416)^2/2 + m(m-1)(m-2)(0.0416)^3/6 + \cdots$$

$$\cong \sum_1^m \rho_i + m(m-1)(0.000,866) + m(m-1)(m-2)(0.000,0120). \tag{18}$$

More terms must be added to (18) unless condition (7)(d), that $m \leqq 20$, is satisfied, but it usually is in practice, if the other conditions of (7) are. In equation (4) we now write $n! \cong n^n e^{-n}(2\pi n)^{1/2}$ with a maximum relative error of $1 - e^{-1/(12n)} \cong 1/(12n)$, so that, finally,

$$P_{k_1 \cdots k_m} \cong \phi(1 + a_1) \cdots (1 + a_m), \tag{19}$$

where

$$\phi = \phi_0 e^{-\chi^2/2}, \quad \phi_0 = (2\pi n)^{(1-m)/2}(p_1 \cdots p_m)^{-1/2}, \quad \chi^2 = \sum_1^m k_i^2/np_i;$$

and the maximum relative error in P is numerically less than

$$\gamma = \sum_1^m \rho_i + m(m-1)(0.000,866) + m(m-1)(m-2)(0.000,012) + 1/(12n). \tag{20}$$

Up to this point we have derived in (19) an approximation to the ordinate of the point binomial solid at the point $(k_1, \cdots, k_{m-1})$. When $m=2$ this reduces to Laplace's well known form for the point binomial. What we really seek however is an approximation to the sum of such ordinates. We shall accordingly find the sum of all $P_{k_1 \cdots k_m}$'s for all values of the k's numerically less than a prescribed fixed set. This prescribed set we shall now call $(k_1, \cdots, k_m)$, using $(k_1', \cdots, k_m')$ for the sets of smaller numbers. It is necessary to explain in more detail precisely what ordinates, and parts of ordinates, our sum is to include. The case of the point binomial shall serve as our model. Putting $m=2$ for a moment we have

$$(p_1 + p_2)^n = p_1^n + \cdots + {}_nC_t p_1^t p_2^t + \cdots + p_2^n = u_0 + \cdots + u_t + \cdots + u_n.$$

Let us think of a histogram in which the base of each rectangle is unity and the altitude one of the u's. If it happens that the mean t, $\bar{t} = p_2 n$, is the coördinate of the middle or one end of one of these bases, then when we add a positive number k so that we come to the middle of a base on the right of the mean, it will also happen that when we subtract k we shall arrive at the middle of a base on the left. Then if we add all the ordinates at these two

end points and at the intermediate points we shall have the same result as if we added the corresponding rectangles of the histogram, i.e., we need to add those rectangles which lie in the interval $(\bar{\imath}-k-1/2,\ \bar{\imath}+k+1/2)$. But, if the mean is not so situated, then when we add k so as to arrive at the middle of a base on the right it will happen that when we subtract k we shall not arrive at the middle of a base on the left. The sum desired in this case is defined to be the total area of those rectangles or parts of rectangles included in the same interval as before. Now, in our general case when $m \neq 2$ each base of unit length is replaced by a unit cube of $m-1$ dimensions, and the sum desired is the sum of volumes of parallelepipeds found by multiplying the volume of each of these cubes by the length of the ordinate P supposed erected at the middle point of the cube. In case $p_1 n_1$, $p_2 n_2$, etc., are integers, as happens in our most important application, then whole parallelepipeds only are included in our sum. If not, partial parallelepipeds may be included; the summation in any event shall extend over the intervals

$$(p_i n - k - \tfrac{1}{2},\ p_i n + k + \tfrac{1}{2}) \qquad (i = 1, \cdots, m-1).$$

We now note from (11) that, if k_i, $i \leqq m-1$, is replaced by $-k_i$, a_{k_i} becomes $-a_{k_i}$, but ϕ is unchanged. To obtain then the approximation corresponding to (19) for any of the points whose coördinates, referred to the general mean of the solid, are $(\pm k_1, \cdots, \pm k_{m-1})$, we have only to change the signs of certain of the a_{k_i}'s, $i \leqq m-1$, and then obtain a_{k_m} from the formula

$$a_{k_m} = \frac{k_m}{6np_m}\left(\frac{k_m^2}{np_m} - 3\right), \quad k_m = -(k_1 + \cdots + k_{m-1});$$

but we must be careful to insist here that both np_i and k_i shall be integers, for by (2) $np_i + k_i$ was an integer and so now $np_i - k_i$ would have to be. In case these numbers are not integers we shall still assume that the formal approximation holds for those ordinates which are at points having integral coördinates and lying in unit cubes contiguous to the one enclosing the point $(\pm k_1, \cdots, \pm k_{m-1})$; for if n is fairly large the relative errors involved at two such adjacent ordinates are almost exactly the same.

In forming our sum, we begin by adding the P's at all the 2^{m-1} vertices of the outer parallelepiped, i.e., at the points $(\pm k_1, \cdots, \pm k_{m-1})$, and denote this sum by $\overline{P}_{k_1 \cdots k_m}$. The coefficient of ϕ in (19) is

$$(21) \qquad 1 + \sum a_i + \sum a_i a_j + \sum a_i a_j a_m + \cdots + (a_1 \cdots a_m), \quad i \neq j \neq k, \text{ etc.}$$

When we form our sum for all the vertices mentioned, certain of the terms that appear like those of (21) will cancel in pairs. To prove this, we note first that, since $k_m = -(k_1 + \cdots + k_{m-1})$, if we change all the signs of

$k_1, k_2, \cdots, k_{m-1}$, we shall change the sign only of k_m, and hence the sign only of a_{k_m}; but that, if we change some only of the signs of $k_1, k_2, \cdots, k_{m-1}$, we may change the numerical value of k_m and hence of a_{k_m}. It follows that by pairing the vertices properly the terms of the type $\sum a_i$ will vanish, also, if m is odd, terms of the type* $(a_1 \cdots a_m)$. To see that this last group vanishes, we note that, if $(k_1, \cdots, k_{m-1})$ is a specified vertex, these k's being positive or negative, there exists also a vertex $(-k_1, \cdots, -k_{m-1})$. For the first vertex our term involves $a_1 \cdots a_m$, and for the second it involves $(-a_1) \cdots (-a_m)$, which equals $-a_1 \cdots a_m$, if m is odd. By analogous reasoning it may be shown now that, by proper pairing, all those terms of the other types involved in (21) will cancel in pairs, except the first, and except perhaps those that contain a_m. Now in $\sum a_i a_j$ there are ${}_{m-1}C_1$ terms that involve a_m, in $\sum a_i a_j a_k$ there are ${}_{m-2}C_2$ such terms, etc. Each of these must be counted 2^{m-1} times. So, finally,

$$\overline{P}_{k_1 \cdots k_m} \leqq (2^{m-1})(1 + 0 + {}_{m-1}C_1 \max|a_m a_i| + {}_{m-2}C_2 \max|a_m a_i a_j| + \cdots)\phi$$

$$\leqq (2^{m-1})\left(1 + \frac{m-1}{100} + \frac{(m-1)(m-2)}{2 \cdot 1000} + \frac{(m-1)(m-2)(m-3)}{3 \cdot 2 \cdot 10{,}000} + \cdots\right)\phi$$

and so

$$\overline{P}_{k_1 \cdots k_m} \cong 2^{m-1}\phi, \tag{22}$$

with a maximum absolute error in $\overline{P}$ of

$$\eta = 2^{m-1}\phi\left(\gamma + \frac{m-1}{100} + \frac{(m-1)(m-2)}{2 \cdot 1000} + \frac{(m-1)(m-2)(m-3)}{3 \cdot 2 \cdot 10{,}000} + \cdots\right)$$

$$\leqq 2^{m-1}\phi\left(\gamma + \frac{m-1}{100} + \frac{(m-1)(m-2)(1.86)}{2000}\right),$$

and therefore

$$\eta \leqq 2^{m-1}\phi\{\gamma + (m-1)/100 + (m-1)(m-2)/1000\}. \tag{23}$$

What we have just done for the parallelepiped $k_1 \cdots k_m$ we may repeat for each of the smaller parallelepipeds $k_1' \cdots k_m'$, obtaining a sum in each case which we may denote by $\overline{P}_{k_1' \cdots k_m'}$. Then by adding all these $\overline{P}$'s together we shall obtain the sum originally sought, viz.:

$$Q = \sum \phi_0 e^{-\chi^2/2}. \tag{24}$$

This is Pearson's expression which he has evaluated by means of a definite integral and tabulated,† but his $(1-P)$ equals my Q, and his n' equals my m.

* This fact is important only when $m=3$.

† Tables, loc. cit., pp. 26 ff.

In order to obtain an estimate of the error in Q we might use the fact that the absolute error in $\overline{P}_{k_1'\cdots k_m'}$ is no greater, *relatively*, than in $\overline{P}_{k_1\cdots k_m}$, but this, although true, would lead us to an inequality too wide to be useful, and so it is better to make an assumption. We shall assume that the *absolute* error in $\overline{P}_{k_1'\cdots k_m'}$ is not greater than in $\overline{P}_{k_1\cdots k_m}$. This is a reasonable assumption, and we know that when the k's are nearly zero, the error also is nearly zero. However, it must be used with caution, for if the k's were allowed to be quite large the value of ϕ would be so exceedingly small that the error η would also be exceedingly small, no matter what the other factors of η might be, and we could not be sure that this absolute error would not be exceeded at points nearer the general mean of the solid; but, as the k's have been considerably restricted by (7), we are not in serious danger on account of that possibility. We shall keep in mind the fact that from now on we are only estimating our maximum error instead of computing it by using the phrase "estimated maximum error." The estimated maximum absolute error in Q as approximated by (23) is $E=\eta(|k_1|+\frac{1}{2})\cdots(|k_{m-1}|+\frac{1}{2})$; or, in a form suitable for computation:

$$
\begin{aligned}
&E = \theta\phi\tau,\ \theta = (|k_1| + \tfrac{1}{2})\cdots(|k_{m-1}| + \tfrac{1}{2})2^{m-1},\\
&\phi = e^{-\chi^2/2}/D,\ D^2 = (np_1)\cdots(np_{m-1})(p_m)(2\pi)^{m-1},\ \chi^2 = \sum_1^m k_i^2/(np_i),\\
&\tau = \sum_1^m \rho_i + (m-1)(0.000,866m + 0.01)\\
&\qquad + (m-1)(m-2)(0.000,012m + 0.001),\\
&\sum_1^m \rho_i = 0.094\sum_1^m (np_i - |k_i|)^{-1} + 0.425\sum_1^m (k_i/(np_i))^2\\
&\qquad + 0.332\sum_1^m |k_i/(np_i)|^3 + 0.176\sum_1^m (k_i/(np_i))^4.
\end{aligned}
\tag{25}
$$

Sometimes it is quite sufficient to know that always $\tau<1.55$. If this gives a sufficiently small value of E, the computation is greatly shortened. Finally, it is necessary to notice that the conditions (7) need not be satisfied exactly in order that the above expression for E may hold approximately, small changes in the conditions (except in (7)(d)) producing small changes in E; but we must not depart from them widely.

5. Numerical examples

The χ-test may now be used to establish rigorously statements which heretofore one could not be truly sure of. As an example we choose the first one used by Pearson in which he appeared to show a bias in dice.* I

* Philosophical Magazine, loc. cit., p. 167.

give below Pearson's figures and also a set of revised figures. My conditions (7) were not nearly satisfied in his last two groups, and so in my proposed revision I have lumped together his last three groups. This of course amounts to a restatement of the whole problem, but the new problem thus created will serve the general purpose Pearson had in view substantially as well as the old one. In this new problem all my conditions are satisfied except for the figures in parentheses, and in these cases they are nearly satisfied, and so, as observed at the close of §4, we may be sure that our E is approximately correct. The second column, marked f, consists of the "observed frequency of dice with five or six points when a cast of twelve dice was made 26,306 times." The column marked np (Pearson's m) consists of the corresponding theoretical frequencies. The first column, marked "dice", indicates the number of dice in a cast which showed 5 or 6 points. Pearson found that $1-Q$ was 0.000,016. This was the probability that a set of dice with no bias would give results as far removed from the theoretical as those observed. It is so small as to make one practically certain that there was bias. But, since in this case we are not sure that it is proper to compute $1-Q$ by the use of the normal function, as he did, I have modified the problem slightly so as to be sure. I find $1-Q=0.000,086$, and, although this is not so small as Pearson's value, it is still small enough to be convincing.

Pearson's Problem				Author's Revision				
Dice	f	np	k	f	k	$np-\|k\|$	ω	α
0	185	203	−18				.0887	1.596
1	1149	1217	−68				.0559	3.800
2	3265	3345	−80				.0239	1.913
3	5475	5576	−101				.0181	1.829
4	6114	6273	−159				.0253	4.030
5	5194	5018	176		No Change		.0351	6.173
6	3067	2927	140				.0478	(6.696)
7	1331	1254	77				.0614	4.728
8	403	392	11				.0281	0.309
9	105	87	18				(.2069)	3.724
10	14	13	1					
11	4	1	3	18	4	10	(.2857)	1.143
12	0	0	0					
$x^2=43.872$, $m=13$, $1-Q=0.000,016$.				$x^2=35.941$, $m=11$, $1-Q=0.000,086$. By (25) $E=0.000,004$; by the shortened form of (25), taking $\tau<1.55$, $E<0.000,017$.				

It is not always feasible to rephrase a problem in this fashion and still keep the general purpose of the investigation in view. Sometimes, however, another sort of revision is possible, which may as well be illustrated in a further, though in this case needless, refinement of the example above. In the revision of that example the last value of ω was 0.2857, instead of the 0.2 permitted by (7), and accordingly we had to think of E and therefore of $1-Q$ as possibly a little larger than as stated. Let us now decrease this ω to less than 0.2 by arbitrarily changing the observed frequency opposite it from 18 to 16 and compensating for this by adding arbitrarily 2 to the frequency 6144, where np is large and k is negative. The total effect is again to set up a new problem: Pearson's was the first, our revision a second, and now we have a third; but, since no k has been numerically increased in passing from the second to the third, the value of Q for the third is known to be smaller than the value of Q sought in the second, since we would now be adding up frequencies over a parallelepiped enclosed in the one used before. We should thus arrive rigorously at a lower limit for the second Q, i.e., at an upper limit for the second $1-Q$, and this might still be small enough to support the general conclusion that there was bias in the dice. We do here in fact obtain $\chi^2=34.983$, $1-Q=0.000,13$, and so we are sure that the value of $1-Q$ in the second problem was less than 0.000,13. E has very nearly the same value as before.

WESLEYAN UNIVERSITY,
MIDDLETOWN, CONN.

THE SINGULAR POINTS OF ANALYTIC SPACE-CURVES*

BY
ARTHUR RANUM

1. The singular points of analytic space-curves have been investigated by various writers.† The object of this paper is to simplify the theory and also to make it self-dual by introducing the dual counterparts of the ordinary curvature and spherical curvature. Let

$$x = x(t), \quad y = y(t), \quad z = z(t) \tag{1}$$

be the equations of a twisted curve, the functions being analytic. Then in the neighborhood of any given point P_0 of the curve, at which $t=t_0$, there exists a certain region of the complex t-plane, within which x, y, z are expressible as series of powers of $t-t_0$ (with positive integral exponents), and such that to different values of t in the region will correspond different points.

The point P_0 will be *singular*, in the broader sense which we shall adopt,‡ if

$$\begin{vmatrix} x' & y' & z' \\ x'' & y'' & z'' \\ x''' & y''' & z''' \end{vmatrix} = 0, \tag{2}$$

when $t=t_0$; here $x'=dx/dt$, etc. The points for which $x'=y'=z'=0$, to which the term singular is sometimes confined, are therefore only a special case.

2. With every point P_0, singular or non-singular, there is associated a triad of positive integers (α, β, γ), such that if we put

---

* Presented, in part, to the American Mathematical Society, December 27, 1915. Abstract in the Bulletin of the American Mathematical Society, vol. 22 (1916), pp. 265, 268.

† Fine, American Journal of Mathematics, vol. 8 (1886), pp. 156–177; Björling, Archiv der Mathematik und Physik, (2), vol. 8 (1890), pp. 83–91; Wölffing, Archiv der Mathematik und Physik, (2), vol. 15 (1897), pp. 146–158; Burali-Forti, Atti di Torino, vol. 36 (1901), pp. 935–938; Mehmke, Zeitschrift für Mathematik und Physik, vol. 49 (1903), pp. 62–83; Saurel, Annals of Mathematics, (2), vol. 7 (1905), pp. 3–9; von Lilienthal, *Vorlesungen über Differentialgeometrie*, vol. 1, Leipzig, 1908, pp. 242–262; Meder, Crelle's Journal, vol. 137 (1910), pp. 83–144; Zindler, Wiener Sitzungsberichte, vol. 127I (1918), pp. 871–918; Study, Mathematische Annalen, vol. 87 (1922), pp. 207–228.

‡ The definitions used by Lilienthal (loc. cit., p. 182), Study (p. 212), Burali-Forti (p. 935) and Mehmke (p. 71) are all practically the same as this; but no writer, so far as I am aware, writes the condition in this simple form. Lilienthal gives three possibilities, but the first two are included in the third.

$$(3) \qquad \Delta = \Delta(\alpha,\beta,\gamma) = \begin{vmatrix} x^{(\alpha)} & y^{(\alpha)} & z^{(\alpha)} \\ x^{(\alpha+\beta)} & y^{(\alpha+\beta)} & z^{(\alpha+\beta)} \\ x^{(\alpha+\beta+\gamma)} & y^{(\alpha+\beta+\gamma)} & z^{(\alpha+\beta+\gamma)} \end{vmatrix}$$

then

$$(3') \qquad \Delta \neq 0, \quad \text{when} \quad t = t_0,$$

whereas every determinant $\Delta(\alpha', \beta', \gamma')$ vanishes, if the triad $(\alpha', \beta', \gamma')$ precedes (α, β, γ) lexicographically, which means that either

$$\alpha' < \alpha,$$

or

$$\alpha' = \alpha, \quad \beta' < \beta,$$

or

$$\alpha' = \alpha, \quad \beta' = \beta, \quad \gamma' < \gamma.^*$$

Hence P_0 is a non-singular point, if $\alpha=\beta=\gamma=1$, and is singular, if $\alpha+\beta+\gamma > 3$. We shall call (α, β, γ) the *type*† of the point P_0 or of the curve (namely the region in the neighborhood of P_0), writing $P_0=(\alpha, \beta, \gamma)$.

By the use of homogeneous coördinates it has been shown‡ that the type of a curve is invariant under the group of collineations, and that every correlation transforms a curve of type (α, β, γ) into one of type (γ, β, α), so that these two types are dual. Hence the type of a curve is a concept belonging to projective, as well as to metric, differential geometry.

We define w and W by the equations

$$(4) \qquad \begin{aligned} w &= \Big(\sum (x^{(\alpha)})^2\Big)^{1/2}, \\ W &= \Big(\sum (y^{(\alpha)} z^{(\alpha+\beta)} - z^{(\alpha)} y^{(\alpha+\beta)})^2\Big)^{1/2}, \end{aligned}$$

where the summations are cyclic with respect to x, y, z, and where t is put equal to t_0.

3. From now on we shall assume that t is real and that x, y, z are real functions of t. Then the condition (3′) shows that $wW \neq 0$; and by choosing the positive square roots we have $w>0$, $W>0$.

If we substitute a new parameter t for $t-t_0$ and also choose the coördinate axes so that the origin O coincides with P_0, the x-axis with the tangent at P_0, and the xy-plane with the osculator (osculating plane) at P_0, then the equations of the curve (in the given region or interval) become

---

* See Lilienthal, pp. 242–251; Mehmke, pp. 68–70; Study, p. 210.

† Lilienthal appears to be the only writer who uses the integers α, β, γ, which he calls ν_1, ν_2, ν_3; Study calls the triad $(\alpha-1, \beta-1, \gamma-1)$ the *characteristic* of the point P_0, writing it (k_1, k_2, k_3).

‡ Björling, p. 85; Study, p. 212.

$$x = at^{\alpha} + \cdots,$$

(5) $$y = bt^{\alpha+\beta} + \cdots,$$

$$z = ct^{\alpha+\beta+\gamma} + \cdots,$$

where $abc \neq 0$ and we can choose $a>0$, $b>0$, $c>0$. Then

(6) $$w = \alpha!a, \quad W = \alpha!(\alpha+\beta)!ab, \quad |\Delta| = \alpha!(\alpha+\beta)!(\alpha+\beta+\gamma)!abc,$$

and therefore

(7) $$a = \frac{w}{\alpha!}, \quad b = \frac{W}{w(\alpha+\beta)!}, \quad c = \frac{|\Delta|}{W(\alpha+\beta+\gamma)!}.$$

The exponents of t in the terms actually present in the expansions (5) will have no common factor; but of course α, β, γ may have a common factor.

The positive branch of the curve, corresponding to small positive values of t, will lie in the octant $[+1, +1, +1]$, in which x, y, z are all positive; and the negative branch will lie in the octant $[(-1)^{\alpha}, (-1)^{\alpha+\beta}, (-1)^{\alpha+\beta+\gamma}]$, which is therefore determined by the evenness or oddness of α, β, and γ. This gives the first rough classification of singular points into eight *categories*.

We also arrive at the same classification in a more significant, because self-dual, manner by considering the curve to be generated by three *elements*, namely a point, a line (the tangent), and a plane (the osculator), and allowing each of these elements to be either *ordinary* or a *reversal*-element (Rückkehrelement).

It is known* that as t decreases (or increases) through the value zero, then at the origin O:

the generating point continues or reverses the direction of its motion along the tangent, according as α is odd or even;

the generating osculator continues or reverses the direction of its rotation about the tangent, according as γ is odd or even;

and the generating tangent continues or reverses the direction of its rotation in the osculator (or about the point), according as β is odd or even.

4. These results are expressed in the following table. In each case, as t decreases through the value zero, the generating point emerges from the first octant $[+1, +1, +1]$ into the octant indicated, whose number, given in the last column, is also the number of the category to which the curve (or point) belongs. The octants 1, 2, 3, 4, for which $z>0$, are numbered like the corresponding quadrants in the xy-plane, and the octants 5, 6, 7, 8 are

* Fine, pp. 166, 173; Saurel, pp. 6–7.

symmetric to the respective octants 1, 2, 3, 4 with respect to the origin, and *not* with respect to the xy-plane, as Lilienthal (p. 202) numbers them.

Type α	Type β	Type γ	Octant ($t<0$) $(-1)^{\alpha}$	$(-1)^{\alpha+\beta}$	$(-1)^{\alpha+\beta+\gamma}$	Number of Octant and of Category
even	even	even	+1,	+1,	+1	1
odd	odd	even	−1,	+1,	+1	2
odd	even	odd	−1,	−1,	+1	3
even	odd	odd	+1,	−1,	+1	4
odd	even	even	−1,	−1,	−1	5
even	odd	even	+1,	−1,	−1	6
even	even	odd	+1,	+1,	−1	7
odd	odd	odd	−1,	+1,	−1	8

In the case of a singular point of the 2d, 3d, or 4th category, one of the three generating elements is a reversal-element and the other two are ordinary; the 2d and 4th categories are dual and the 3d is self-dual. Similarly the 5th and 7th are dual and the 6th is self-dual. The 1st and 8th are also self-dual; the 8th includes the non-singular points (1, 1, 1).*

The positive and negative branches of the curve will lie on the same side of the osculating plane (categories 1, 2, 3, 4) or on opposite sides (categories 5, 6, 7, 8), according as the number of reversal-elements is odd or even.

5. The interval $(-h<t<h)$ within which the curve is defined by (5) can be chosen such that every point except the origin O is non-singular. At any such point $P=(x, y, z)$ consider the six variables

$$s \quad \rho \quad \theta \quad \tau \quad \sigma \quad \eta$$

where $1/\rho$ and $1/\tau$ are the ordinary curvature and torsion and where

$$s=\int_0^t (\sum x'^2)^{1/2}dt, \qquad \theta=\int_0^t ds/\rho, \qquad \eta=\int_0^t ds/\tau \quad \text{and} \quad \sigma=\rho/\tau.$$

Then $\rho=ds/d\theta$, $\tau=ds/d\eta$ and $\sigma=d\eta/d\theta$; ds, $d\theta$ and $d\eta$ are the element of arc,

* Mehmke (pp. 65–67) and Zindler (pp. 876–879) give distinctive names to these eight categories; Zindler's names seem to be highly appropriate, except that in accordance with his definition a point of the 8th category is called non-singular (gewöhnlich); it would be better to call it a "Gegenhenkel."

the angle of contingence and the angle of torsion, respectively; s is dual to η, ρ to σ, and θ, τ are self-dual. Following Plucker* we shall call σ the *conical curvature* at P.

We define the value of ρ at a singular point O to be the limit of its value at P as $t \to 0$, including the case where $\rho \to \infty$; similarly for τ and σ.

6. We now proceed to find a second classification of the singular points O into thirteen *classes*, depending on whether the values of ρ, τ and σ at O are zero, finite and different from zero, or infinite. Since $\sigma = -\cot \omega$, where ω is the angle between the tangent and the rectifying line, σ will be zero, infinite, or neither, according as the rectifying line coincides with the binormal, with the tangent, or with neither one.

From (5) by means of known formulas we easily derive for the values of ρ, τ and σ at the non-singular point P the following expansions in ascending integral powers of t, in which the exponents of t in the initial terms are sometimes negative:

$$\rho = \frac{\alpha^2 a^2}{\beta(\alpha+\beta)b} t^{\alpha-\beta} + \cdots,$$

$$\tau = -\frac{\alpha\beta(\alpha+\beta)ab}{\gamma(\beta+\gamma)(\alpha+\beta+\gamma)c} t^{\alpha-\gamma} + \cdots, \tag{8}$$

$$\sigma = -\frac{\alpha\gamma(\beta+\gamma)(\alpha+\beta+\gamma)ac}{\beta^2(\alpha+\beta)^2 b^2} t^{\gamma-\beta} + \cdots.$$

We also find that

$$s' = \alpha a t^{\alpha-1} + \cdots,$$

$$\theta' = \frac{\beta(\alpha+\beta)b}{\alpha a} t^{\beta-1} + \cdots, \tag{9}$$

$$\eta' = -\frac{\gamma(\beta+\gamma)(\alpha+\beta+\gamma)c}{\beta(\alpha+\beta)b} t^{\gamma-1} + \cdots,$$

and therefore that

$$s = a t^{\alpha} + \cdots,$$

$$\theta = \frac{(\alpha+\beta)b}{\alpha a} t^{\beta} + \cdots, \tag{10}$$

$$\eta = -\frac{(\beta+\gamma)(\alpha+\beta+\gamma)c}{\beta(\alpha+\beta)b} t^{\gamma} + \cdots.$$

---

* *Geometrie des Raumes*, 1846, pp. 34–35. See also F. Müller, Wiener Sitzungsberichte, vol. 126 (1917), pp. 311–312. In the Quarterly Journal of Mathematics, vol. 46 (1915), p. 36, I called $\sigma(=1/r')$ the plane-curvature.

The leading coefficient can never be zero in any of these nine expansions. The leading terms in (8_1), (8_2) and (8_3), under the respective hypotheses $\alpha=\beta$, $\alpha=\gamma$ and $\beta=\gamma$, become equal to l, m and n, respectively, where

$$l = \frac{a^2}{2b}, \quad m = -\frac{\beta ab}{(2\alpha+\beta)c}, \tag{11}$$
$$n = -\frac{2\alpha(\alpha+2\beta)ac}{(\alpha+\beta)^2b^2}.$$

If $\alpha=\beta=\gamma$, then m and n become simpler in form and we have

$$l = a^2/(2b), \quad m = -ab/(3c), \quad n = -3ac/(2b^2).$$

We now pass to the limit by letting $t\to 0$ and obtain for the values of ρ, τ and σ at the origin O the following:

$$\rho = \left\{\begin{matrix}0\\ l\\ \infty\end{matrix}\right\} \text{ according as } \alpha \left\{\begin{matrix}>\\ =\\ <\end{matrix}\right\} \beta,$$
$$\tau = \left\{\begin{matrix}0\\ m\\ \infty\end{matrix}\right\} \text{ according as } \alpha \left\{\begin{matrix}>\\ =\\ <\end{matrix}\right\} \gamma, \tag{12}$$
$$\sigma = \left\{\begin{matrix}0\\ n\\ \infty\end{matrix}\right\} \text{ according as } \gamma \left\{\begin{matrix}>\\ =\\ <\end{matrix}\right\} \beta,$$

where l, m and n, given by (11), are finite and $\neq 0$. By means of (7), (3) and (4) we can, if we wish, express l, m and n in terms of the original equation (1) of the curve.

By examining (12_2) and (11_2) we see that if a point O is of type (α, β, α), then the absolute value of τ at O is always less than ab/c, and is *greater than, equal to, or less than* $ab/(3c)$, *according as* β *is greater than, equal to, or less than* α; also if α is fixed and $\beta\to\infty$, then $|\tau|\to ab/c$, and if β is fixed and $\alpha\to\infty$, then $|\tau|\to 0$.

By a similar examination of (12_3) and (11_3) we see that if a point is of type (α, β, β), then $|\sigma|$ is always less than $2ac/b^2$, and is $>$, $=$, or $<3ac/(2b^2)$, according as β is $<$, $=$, or $>\alpha$. Hence ω (an acute angle) is always greater than $\cot^{-1}(2ac/b^2)$, and is $<$, $=$, or $>\cot^{-1}(3ac/(2b^2))$, according as β is $<$, $=$, or $>\alpha$; also if $\alpha\to\infty$, then $\omega\to\cot^{-1}(2ac/b^2)$, and if $\beta\to\infty$, then $\omega\to\pi/2$.

7. Formulas (12) show us that the class to which any real and finite point of type (α, β, γ) belongs depends only on the relative magnitudes of α, β, and γ. From this fact or the relation $\rho=\sigma\tau$ we see that the number of classes, instead of being 3^3, reduces to 13, as is shown in the following table:

Class-Number	Type	ρ	τ	σ
1	$\alpha=\beta=\gamma$	$\frac{a^2}{2b}$	$-\frac{ab}{3c}$	$-\frac{3ac}{2b^2}$
2	$\alpha=\gamma<\beta$	∞	m	∞
3	$\alpha=\gamma>\beta$	0	m	0
4	$\alpha=\beta<\gamma$	l	∞	0
5	$\gamma=\beta<\alpha$	0	0	n
6	$\alpha=\beta>\gamma$	l	0	∞
7	$\gamma=\beta>\alpha$	∞	∞	n
8	$\alpha<\beta<\gamma$	∞	∞	0
9	$\gamma<\beta<\alpha$	0	0	∞
10	$\alpha<\gamma<\beta$	∞	∞	∞
11	$\gamma<\alpha<\beta$	∞	0	∞
12	$\beta<\alpha<\gamma$	0	∞	0
13	$\beta<\gamma<\alpha$	0	0	0

In each of the six classes 2–7 exactly two of the type-integers α, β, γ are equal and therefore exactly one of the variables ρ, τ, σ is finite and $\neq 0$. In each of the last six classes 8–13 the three integers are all distinct and ρ, τ, σ are all $=0$ or ∞. Classes 1, 2 and 3, in which τ is finite and $\neq 0$, are self-dual; and 4 is dual to 5, 6 to 7, 8 to 9, 10 to 11, and 12 to 13. In any pair of dual classes the values of ρ and σ in one are interchanged in the other (except for the difference between l and n).

8. From our cross-classification of singular points into eight categories and thirteen classes, by combination, we obtain a classification into *species*; the species of a point depends, therefore, on the evenness or oddness of α, β, γ and also on their relative magnitudes. Since these two properties are not entirely independent, the number of species is less than $8 \cdot 13 = 104$.

First, a singular point of class 1 must clearly be of category 1 or 8. Hence *if ρ, τ and σ are all finite and different from zero, the curve must either emerge (for $t<0$) in the 8th octant or return to the 1st octant.* This gives two species.

Next, in each of the six classes 2–7 two of the integers α, β, γ, being equal, must be either both odd or both even, while the third may be odd or even. This gives $6 \cdot 4 = 24$ species.

In particular, putting $\alpha=\beta$ ($\rho=l$), we see that *the only categories of singular points at which the ordinary curvature $1/\rho$ can be finite and different from zero are the 1st, 8th, 2d and 7th, for which the negative branch of the curve lies in one of the octants* $[(-1)^\alpha, +1, (-1)^\gamma]$.

Similarly ($\alpha=\gamma$, $\tau=m$) *the only categories of points at which the torsion $1/\tau$ can be finite and different from zero are the 1st, 8th, 3d and 6th, for which*

the negative branch of the curve lies in one of the octants $[(-1)^{\alpha}, (-1)^{\alpha+\beta}, (-1)^{\beta}]$. Of these four octants no two are adjacent (symmetric with respect to a plane).

Again ($\beta=\gamma$, $\sigma=n$) *the only categories of points at which the conical curvature* σ *can be finite and different from zero are the 1st, 8th, 4th, and 5th, for which the negative branch of the curve lies in one of the octants* $[(-1)^{\alpha}, (-1)^{\alpha+\beta}, (-1)^{\alpha}]$.*

In other words, *if the negative branch of the curve lies in the 2d or 7th octant,* τ *and* σ *must be zero or infinite; if in the 3d or 6th octant,* ρ *and* σ *must be zero or infinite; if in the 4th or 5th octant,* ρ *and* τ *must be zero or infinite.*

Finally, in each of the six classes 8–13 the integers α, β, γ are unrestricted as to oddness or evenness; that is, a point at which ρ, τ, σ are all zero or infinite may belong to any one of the eight categories. This gives $6 \cdot 8 = 48$ species. *The total number of species is therefore* $2+24+48=74$.†

9. Since every point belonging to a self-dual class ($\alpha=\gamma$, $\tau=m$) will also belong to a self-dual category, the self-dual species are precisely those that belong to the classes 1, 2, 3, and their number is $2+4+4=10$.

It is evident that the species of a point is invariant under all the real collineations that carry the point into a *finite* point.‡

If the curve is imaginary, our entire classification falls to the ground, classes as well as categories; for by (4), (7) and (11) we see that w and W may vanish and that a, b, c, l, m, n may not be determined. For instance, the origin is a non-singular point (1, 1, 1) of the curve $x=it$, $y=t+t^2$, $z=t^3$, and yet since $w=0$, $W=2i$ and $\Delta=12i$, we have $\rho=\sigma=0$, $\tau=i/3$, which contradicts (12).

The simplest singular points are (2, 1, 1), (1, 2, 1) and (1, 1, 2); the first is said to be a *stationary* point (of the first degree), at the second the tangent is said to be stationary, and at the third the osculator is said to be stationary. Fine (p. 174) has shown that if the curve is algebraic, then with respect to its effect on the order, class, rank, and genus of the curve, a singular point of type (α, β, γ) is equivalent to $\alpha-1$ stationary points, $\beta-1$ stationary tangents and $\gamma-1$ stationary osculators.

Study (p. 214) has also shown that (α, β, γ) is an α-fold point of the curve, the corresponding tangent is β-fold, and the osculator is γ-fold.

---

* The first two parts of this triple theorem were proved by Mehmke (pp. 75, 77).

† Burali-Forti (p. 938) and Mehmke (p. 78), by ignoring σ, obtain a classification that is not self-dual. Instead of 13 classes and 74 species they obtain only 9 classes and 50 species. For them our classes 5, 9 and 13 coincide; also 7, 8 and 10. For them the points (1, 2, 4), (1, 2, 2) and (1, 4, 2) are of the same species, although the corresponding values of σ are 0, n and ∞, respectively.

‡ Meder (loc. cit.) studies the "Ordnungszahlen" of ρ, τ and various other variables at an *infinite* point.

PROJECTIONS ON THE COÖRDINATE PLANES

10. Now consider the orthogonal projections of a curve C on the osculating, rectifying and normal planes at a point $O=(\alpha, \beta, \gamma)$. From (5) we see that the types of these three plane curves at their common point are

$$O_{xy} = (\alpha,\beta), \qquad (13)$$

$$O_{xz} = (\alpha,\beta+\gamma),$$

$$O_{yz} = (\alpha+\beta,\gamma),$$

respectively. Also consider a curve C' dual to C and on it the point O' that corresponds to O in a correlation; then $O'=(\gamma, \beta, \alpha)$, $O_{xz}'=(\gamma, \alpha+\beta)$ and $O_{yz}'=(\beta+\gamma, \alpha)$. But in plane geometry a point (λ, μ) is dual to a point (μ, λ). Hence O_{yz}' is dual to O_{xz} and O_{xz}' to O_{yz}. That is, if two space-curves are dual, the projection of either on the *normal plane* at a point O is dual (planar duality) to the projection of the other on the *rectifying plane* at the corresponding point O'.

If $\alpha=\beta+\gamma$, O_{xz} (and therefore also O_{yz}') is self-dual, and conversely. If $\alpha+\beta=\gamma$, O_{yz} (and therefore also O_{xz}') is self-dual, and conversely.

If $\alpha=\gamma$, so that O is self-dual (spatial duality), then $O_{xz}=(\alpha, \alpha+\beta)$ and $O_{yz}=(\alpha+\beta, \alpha)$, which are therefore dual (planar duality), and conversely. That is, *if the torsion at a point O is finite and different from zero, the projections of the curve on the normal and rectifying planes at O are dual, and conversely.*

11. It is known* that the singular points of plane curves are divided into four categories (quadrants), three classes ($\rho=0$, ∞, or neither) and ten species. Consider the possible combinations of categories (classes, species) of the three projections (13). The combinations of categories are obviously eight in number, and there is a one-to-one correspondence between the eight octants in which the negative branch of the curve C may lie and the eight combinations of quadrants in which the negative branches of the projections may lie.†

12. When we come to the combinations of classes or species, we find a very different situation. Let ρ_{xy}, ρ_{xz} and ρ_{yz} be the respective radii of curvature of the three projections (13) at their common point. Of course $\rho_{xy}=\rho$, and its value is determined by (12_1) and (11_1). Similarly ρ_{xz} and ρ_{yz} are determined as follows:

---

* See Burali-Forti, p. 938.

† See the figures drawn by Mehmke, pp. 65–67.

$$\rho_{zz} = \left\{\begin{matrix}0\\k_1\\\infty\end{matrix}\right\}, \text{ according as } \alpha \left\{\begin{matrix}>\\=\\<\end{matrix}\right\} \beta + \gamma,$$

(14)

$$\rho_{yz} = \left\{\begin{matrix}0\\k_2\\\infty\end{matrix}\right\}, \text{ according as } \alpha + \beta \left\{\begin{matrix}>\\=\\<\end{matrix}\right\} \gamma,$$

where

$$k_1 = a^2/(2c) \neq 0, \quad k_2 = b^2/(2c) \neq 0. \tag{15}$$

The number of combinations of classes is only eleven. For if $\alpha \leqq \beta$, then $\alpha < \beta+\gamma$ and γ may be either $<$, $=$, or $> \alpha+\beta$; this gives $2 \cdot 3 = 6$ combinations. But if $\alpha > \beta$, γ may be either $\geqq \alpha+\beta$ or $\leqq \alpha-\beta$, or finally we may have $\alpha+\beta > \gamma > \alpha-\beta$; this gives 5 combinations. Altogether $6+5=11$ combinations, which are exhibited in the following table.

	Type			ρ_{xy} $(=\rho)$	ρ_{zz}	ρ_{yz}
1	$\alpha<\beta$	$\alpha<\beta+\gamma$	$\alpha+\beta<\gamma$	∞	∞	∞
2	"	"	$\alpha+\beta=\gamma$	∞	∞	k_2
3	"	"	$\alpha+\beta>\gamma$	∞	∞	0
4	$\alpha=\beta$	"	$\alpha+\beta<\gamma$	l	∞	∞
5	"	"	$\alpha+\beta=\gamma$	l	∞	k_2
6	"	"	$\alpha+\beta>\gamma$	l	∞	0
7	$\alpha>\beta$	"	$\alpha+\beta<\gamma$	0	∞	∞
8	"	"	$\alpha+\beta=\gamma$	0	∞	k_2
9	"	"	$\alpha+\beta>\gamma$	0	∞	0
10	"	$\alpha=\beta+\gamma$	"	0	k_1	0
11	"	$\alpha>\beta+\gamma$	"	0	0	0

If $\alpha \geqq \beta+\gamma$, then $\alpha > \beta$ and $\alpha+\beta > \gamma$. Hence by (14) and (12_1) or by an inspection of the table, we see that *the only points O for which the radius of curvature ρ_{zz} of the projection of the curve on the rectifying plane at O is not infinite are those for which the radii of curvature ρ_{xy} and ρ_{yz} of the projections on the osculating and normal planes are both zero.*

Putting $\alpha = \beta+\gamma$ or $\alpha+\beta = \gamma$, we easily find, by §4 or by a known theorem on plane curves, that *if the two branches of the curve at a point O lie on opposite sides of the osculating plane, then the radii of curvature ρ_{xz} and ρ_{yz} of its projections on the rectifying and normal planes must be zero or infinite.*

13. The number of combinations of *species* is easily seen to be only 62, although each projection by itself is entirely unrestricted in species (ten cases). There is obviously no one-to-one correspondence between the 13 classes (74 species) of curves and the 11 combinations of classes (62 combinations of species) of their projections. Not only may two curves of different

species have projections belonging to the same combination of species, but two curves of the same species may have projections belonging to different combinations of species (their projections O_{xy} on the osculator must, of course, be of the same species).

The first possibility is illustrated by the types (4, 2, 3) and (4, 2, 5), for which $\tau=0$ and ∞, respectively, whereas their projections are of types (4, 2), (4, 5), (6, 3) and (4, 2), (4, 7), (6, 5), respectively, for which in both cases $\rho_{xy}=0$, $\rho_{xz}=\infty$, $\rho_{yz}=0$. The second possibility is illustrated by (4, 2, 3), just considered, and (6, 2, 3); these two types are clearly of the same species, whereas the projections of the latter type of curve are of types (6, 2), (6, 5), (8, 3), for which $\rho_{xy}=\rho_{xz}=\rho_{yz}=0$.

In spite of this lack of uniformity in the correspondence, there is, of course, a close connection between the values of ρ, τ, σ and the values of $\rho_{xy}(=\rho)$, ρ_{xz}, ρ_{yz}. For instance:

If $\alpha\geqq\beta+\gamma$, then $\alpha>\beta$ and $\alpha>\gamma$. By (14) and (12) this says that if $\rho_{xz}=k_1$ or 0, then $\rho=\tau=0$.

Hence, *for every point O except those at which* $\rho=\tau=0$, *the projection of the curve on the rectifying plane at O has an infinite radius of curvature* ρ_{xz}.

If $\gamma\geqq\beta+\alpha$, then $\gamma>\beta$ and $\gamma>\alpha$. By (14) and (12) this says that if $\rho_{yz}=k_2$ or ∞, then $\sigma=0$ and $\tau=\infty$.

Hence, *for every point O except those at which* $\sigma=0$ *and* $\tau=\infty$, *the projection of the curve on the normal plane at O has a vanishing radius of curvature* ρ_{yz}.

Sections of the tangent surface by the coördinate planes

14. The osculating, rectifying and normal planes at a point $O=(\alpha, \beta, \gamma)$ will cut the tangent surface of the curve in three plane curves $\bar{O}_{xy}$, $\bar{O}_{xz}$ and $\bar{O}_{yz}$, whose character at their common point O we shall now consider. The parametric equations of the tangent surface of the curve (5), §3, are

$$
\begin{aligned}
x &= (at^{\alpha}+\cdots)+u(\alpha at^{\alpha-1}+\cdots),\\
y &= (bt^{\alpha+\beta}+\cdots)+u[(\alpha+\beta)bt^{\alpha+\beta-1}+\cdots],\\
z &= (ct^{\alpha+\beta+\gamma}+\cdots)+u[(\alpha+\beta+\gamma)ct^{\alpha+\beta+\gamma-1}+\cdots],
\end{aligned}
\tag{16}
$$

where t and u are the parameters. The osculator $z=0$ will intersect the surface in points for which

$$u=-\frac{1}{\alpha+\beta+\gamma}t+\cdots.$$

Substituting this expansion of u in (16_1) and (16_2), we find the equations of the section $\bar{O}_{xy}$ to be

$$x = \frac{\beta+\gamma}{\alpha+\beta+\gamma} at^{\alpha} + \cdots, \quad y = \frac{\gamma}{\alpha+\beta+\gamma} bt^{\alpha+\beta} + \cdots. \tag{17}$$

Similarly the equations of $\bar{O}_{xz}$ are found to be

$$x = \frac{\beta}{\alpha+\beta} at^{\alpha} + \cdots, \quad z = -\frac{\gamma}{\alpha+\beta} ct^{\alpha+\beta+\gamma} + \cdots, \tag{18}$$

and the equations of $\bar{O}_{yz}$ to be

$$y = -\frac{\beta}{\alpha} bt^{\alpha+\beta} + \cdots, \quad z = -\frac{\beta+\gamma}{\alpha} ct^{\alpha+\beta+\gamma} + \cdots. \tag{19}$$

Since the leading coefficients in the six expansions of (17), (18) and (19) are all $\neq 0$, the sections $\bar{O}_{xy}$, $\bar{O}_{xz}$ and $\bar{O}_{yz}$ are of the same types as the respective projections O_{xy}, O_{xz} and O_{yz}, as given by (13), §10. Hence everything stated in §§10–13 about the categories, classes and species of the three projections applies equally well to the three sections.

15. Let $\bar{\rho}_{xy}$, $\bar{\rho}_{xz}$ and $\bar{\rho}_{yz}$ be the respective radii of curvature of the three sections at the point O. Their values are easily found to be the following:

$$\bar{\rho}_{xy} = \begin{Bmatrix} 0 \\ a_0 \\ \infty \end{Bmatrix} \text{ according as } \alpha \begin{Bmatrix} > \\ = \\ < \end{Bmatrix} \beta, \tag{20}$$

where

$$a_0 = \frac{(\alpha+\gamma)^2 a^2}{2\gamma(2\alpha+\gamma)b}; \tag{20'}$$

$$\bar{\rho}_{xz} = \begin{Bmatrix} 0 \\ a_1 \\ \infty \end{Bmatrix} \text{ according as } \alpha \begin{Bmatrix} > \\ = \\ < \end{Bmatrix} \beta + \gamma, \tag{21}$$

where

$$a_1 = -\frac{\beta^2 a^2}{2\gamma(2\beta+\gamma)c}; \tag{21'}$$

$$\bar{\rho}_{yz} = \begin{Bmatrix} 0 \\ a_2 \\ \infty \end{Bmatrix} \text{ according as } \alpha+\beta \begin{Bmatrix} > \\ = \\ < \end{Bmatrix} \gamma, \tag{22}$$

where

$$a_2 = -\frac{\beta^2 b^2}{2\alpha(\alpha+2\beta)c}. \tag{22'}$$

Comparing (20) and (20′) with (12_1) and (11_1), we see that when a point is of type (α, α, γ) then

$$\bar{\rho}_{xy} = \frac{(\alpha+\gamma)^2}{\gamma(2\alpha+\gamma)}\rho \neq 0, \tag{23}$$

and therefore

$$\bar{\rho}_{xy} > \rho \,(= \rho_{xy} = a^2/(2b)), \quad \text{and}$$

$$\bar{\rho}_{xy} \left\{ \begin{matrix} > \\ = \\ < \end{matrix} \right\} \frac{4}{3}\rho, \text{ according as } \gamma \left\{ \begin{matrix} < \\ = \\ > \end{matrix} \right\} \alpha. \tag{24}$$

In other words *if the radius of curvature ρ of a curve C at a point O is finite and different from zero, then the radius of curvature $\bar{\rho}_{xy}$ of the section of the tangent surface of C by the osculator at O is always finite and greater than the radius of curvature ρ_{xy} $(=\rho)$ of the projection of C on the osculator, and is greater than, equal to, or less than* $(4/3)\rho$, *according as the radius of torsion τ of C at O is zero, finite and different from zero, or infinite.* The osculating circle of the section and that of the projection lie on the same side of the tangent.

16. By a similar comparison of (21) and (21′) with (14_1) and (15_1), we see that when $\alpha=\beta+\gamma$, then

$$\bar{\rho}_{xz} = -\frac{\beta^2}{\gamma(2\beta+\gamma)}\rho_{xz} \neq 0\,; \tag{25}$$

hence $|\bar{\rho}_{xz}| \neq \rho_{xz}$ and

$$|\bar{\rho}_{xz}| \left\{ \begin{matrix} \geq \\ \leq \end{matrix} \right\} \rho_{xz} \text{ according as } \beta \left\{ \begin{matrix} \geq \\ \leq \end{matrix} \right\} (1+2^{1/2})\gamma\,;$$

in particular if $\beta=\gamma$, $\bar{\rho}_{xz}=-\rho_{xz}/3$.

That is, *at a singular point O of type $(\beta+\gamma, \beta, \gamma)$, the radius of curvature $|\bar{\rho}_{xz}|$ of the section of the tangent surface by the rectifying plane is never equal to, but is always a rational multiple of, the radius of curvature ρ_{xz} of the projection of the curve on the rectifying plane, and is greater or less than ρ_{xz}, according as β is greater or less than* $(1+2^{1/2})\gamma$. The osculating circle of the section and that of the projection lie on opposite sides of the tangent.

In particular, if the point is of type $(2\gamma, \gamma, \gamma)$, then $\bar{\rho}_{xz}=-\rho_{xz}/3$.

Finally, comparing (22) and (22′) with (14_2) and (15_2), we see that if $\alpha+\beta=\gamma$, then

$$\bar{\rho}_{yz} = -\frac{\beta^2}{\alpha(\alpha+2\beta)}\rho_{yz} \neq 0\,; \tag{26}$$

hence $|\bar{\rho}_{yz}| \neq \rho_{yz}$, and

$$|\bar{\rho}_{yz}| \left\{ \begin{matrix} \geq \\ \leq \end{matrix} \right\} \rho_{yz} \text{ according as } \beta \left\{ \begin{matrix} \geq \\ \leq \end{matrix} \right\} (1+2^{1/2})\alpha\,;$$

in particular, if $\beta=\alpha$, $\bar{\rho}_{yz}=-\rho_{yz}/3$.

That is, *at a singular point of type $(\alpha, \beta, \alpha+\beta)$ the radius of curvature $|\bar{\rho}_{yz}|$ of the section by the normal plane is never equal to, but is always a rational*

multiple of, the radius of curvature ρ_{yz} of the projection on the normal plane, and is greater or less than ρ_{yz}, according as β is greater or less than $(1+2^{1/2})\alpha$. The osculating circle of the section and that of the projection lie on opposite sides of their common tangent (the principal normal of the original curve).

In particular, if the point is of type $(\alpha, \alpha, 2\alpha)$, then $\bar{\rho}_{yz} = -\rho_{yz}/3$.

Spherical and quasi-spherical curvature

17. The dual counterpart of the osculating sphere of a twisted curve C at a point O we shall call its quasi-osculating sphere* at O. Its center is the corresponding point on the edge of regression of the rectifying developable and its radius is the distance from that point to the corresponding tangent to C. Since the radius R_1 of the osculating sphere has been called the radius of *spherical curvature* of C, we shall call the radius R_2 of the quasi-osculating sphere the radius of *quasi-spherical curvature* of C. By means of the known formulas

$$R_1^2 = \rho^2 + \left(\frac{\rho'}{\eta'}\right)^2, \qquad R_2 = \left|\frac{s'}{\sigma'}\right|, \tag{27}$$

we proceed to find the values of R_1 and R_2 at a point (α, β, γ).

18. Differentiating (8_1) and (8_3), §6, we obtain

$$\begin{aligned} \rho' &= \frac{\alpha^2(\alpha-\beta)a^2}{\beta(\alpha+\beta)b}t^{\alpha-\beta-1} + \cdots, \\ \sigma' &= \frac{\alpha\gamma(\beta^2-\gamma^2)(\alpha+\beta+\gamma)ac}{\beta^2(\alpha+\beta)^2b^2}t^{\gamma-\beta-1} + \cdots, \end{aligned} \tag{28}$$

where the initial coefficients will vanish under the respective conditions $\alpha=\beta, \beta=\gamma$. Some of the succeeding coefficients may also vanish, and we have the following:

$$\begin{aligned} &\text{if} \quad \alpha = \beta, \quad \rho' = a_i t^i + \cdots \qquad (i \geqq 0,\ a_i \neq 0); \\ &\text{if} \quad \beta = \gamma, \quad \sigma' = b_k t^k + \cdots \qquad (k \geqq 0,\ b_k \neq 0). \end{aligned} \tag{29}$$

From (28_1), (29_1) and (9_3) we see that

$$\begin{aligned} &\text{if} \quad \alpha \neq \beta, \quad \rho'/\eta' = \frac{\alpha^2(\beta-\alpha)a^2}{\gamma(\beta+\gamma)(\alpha+\beta+\gamma)c}t^{\alpha-\beta-\gamma} + \cdots, \\ &\text{if} \quad \alpha = \beta, \quad \rho'/\eta' = -\frac{2\alpha^2 b a_i}{\gamma(\alpha+\gamma)(2\alpha+\gamma)c}t^{i+1-\gamma} + \cdots. \end{aligned} \tag{30}$$

* Quarterly Journal of Mathematics, vol. 46 (1915), pp. 364–366; Mathematische Annalen, vol. 101 (1929), p. 2; G. Loria, *Curve Sghembe Speciali*, vol. 1, 1921, p. 6; Th. Schmid, *Darstellende Geometrie*, vol. 1, 1919, p. 164.

For the sake of brevity we put

$$(31)\qquad g_1 = \left|\frac{\alpha^2(\beta-\alpha)a^2}{\gamma(\beta+\gamma)(\alpha+\beta+\gamma)c}\right|,$$
$$h_1 = \left|\frac{2\alpha^2 b a_i}{\gamma(\alpha+\gamma)(2\alpha+\gamma)c}\right|,$$

so that (30) becomes the following:

$$(32)\qquad \text{if } \alpha \neq \beta, \quad \rho'/\eta' = \pm g_1 t^{\alpha-\beta-\gamma} + \cdots,$$
$$\text{if } \alpha = \beta, \quad \rho'/\eta' = \pm h_1 t^{i+1-\gamma} + \cdots.$$

19. Comparing (27_1), (8_1) and (32_1) we see that if $\alpha \neq \beta$, the significant term in the expression for R_1^2 is $(\rho'/\eta')^2$. Hence

$$(33)\qquad \text{if } \alpha \neq \beta, \ R_1 = \left| g_1 t^{\alpha-\beta-\gamma} + \cdots \right|.$$

On the other hand if $\alpha=\beta$, so that (8_1) reduces to

$$(34)\qquad \rho = a^2/(2b) + \cdots,$$

then by comparing (32_2) and (34) we see that the significant term depends on whether $i+1-\gamma$ is negative, zero or positive. Hence

$$(35)\qquad \text{if } \alpha = \beta \text{ and } \begin{cases} i+1<\gamma, & \text{then } R_1 = \left| h_1 t^{-(\gamma-i-1)} + \cdots \right|, \\ i+1=\gamma, & \text{then } R_1 = (h_1^2 + a^4/(4b^2))^{1/2} + \cdots, \\ i+1>\gamma, & \text{then } R_1 = a^2/(2b) + \cdots. \end{cases}$$

Equations (33) and (35) give the value of R_1 at a non-singular point in the neighborhood of the singular point O. Passing to the limit by letting $t \to 0$, we find for the value of R_1 at O the following cases:

$$(36)\qquad \begin{aligned} &\text{(a)} \quad \text{If } \alpha > \beta+\gamma, \quad R_1 = 0. \\ &\text{(b)} \quad \text{If } \alpha = \beta+\gamma, \quad R_1 = g_1 = a^2/(2c). \\ &\text{(c)} \quad \text{If } \alpha < \beta+\gamma \text{ and } \alpha \neq \beta, \quad R_1 = \infty. \\ &\text{(d)} \quad \text{If } \alpha = \beta \text{ and } i+1 > \gamma, \quad R_1 = a^2/(2b). \\ &\text{(e)} \quad \text{If } \alpha = \beta \text{ and } i+1 = \gamma, \quad R_1 = (h_1^2 + a^4/(4b^2))^{1/2}. \\ &\text{(f)} \quad \text{If } \alpha = \beta \text{ and } i+1 < \gamma, \quad R_1 = \infty. \end{aligned}$$

In cases (b), (d) and (e), in view of (31), R_1 is finite and $\neq 0$. If $\gamma=1$, case (f) does not occur.

20. Turning now to R_2, given by (27_2), we soon find that by using (9_1), (28_2) and (29_2) and putting

$$(37)\qquad g_2 = \left|\frac{\beta^2(\alpha+\beta)^2b^2}{\gamma(\beta^2-\gamma^2)(\alpha+\beta+\gamma)c}\right|, \qquad h_2 = \left|\frac{\alpha a}{b_k}\right|,$$

we obtain the following expansions of R_2:

$$(38)\qquad \text{if } \beta \neq \gamma, \quad R_2 = \left| g_2 t^{\alpha+\beta-\gamma} + \cdots \right| ; \qquad \text{if } \beta = \gamma, \quad R_2 = \left| h_2 t^{\alpha-k-1} + \cdots \right| .$$

Again passing to the limit as $t \to 0$, we find for the value of R_2 at the singular point O the following cases:

$$(39)\qquad \begin{aligned} &\text{(a)} \quad \text{If } \alpha+\beta<\gamma, \quad R_2=\infty . \\ &\text{(b)} \quad \text{If } \alpha+\beta=\gamma, \quad R_2=g_2=\beta^2b^2/(2\alpha(\alpha+2\beta)c). \\ &\text{(c)} \quad \text{If } \alpha+\beta>\gamma, \text{ and } \beta\neq\gamma, \quad R_2=0. \\ &\text{(d)} \quad \text{If } \beta=\gamma \text{ and } k+1>\alpha, \quad R_2=\infty . \\ &\text{(e)} \quad \text{If } \beta=\gamma \text{ and } k+1=\alpha, \quad R_2=h_2. \\ &\text{(f)} \quad \text{If } \beta=\gamma \text{ and } k+1<\alpha, \quad R_2=0. \end{aligned}$$

In cases (b) and (e), R_2 is finite and $\neq 0$. If $\alpha = 1$, case (f) does not occur.

21. From (36) and (39) we easily derive a number of interesting consequences. In view of (14), (21), (22) and (12) we see that R_1 is closely connected with ρ_{zz}, $\bar{\rho}_{zz}$ and ρ, and that R_2 is connected with ρ_{yz}, $\bar{\rho}_{yz}$ and σ. Indeed:

The radius of spherical curvature R_1 at a singular point O is zero, infinite or neither, according as the radii of curvature ρ_{zz} and $\bar{\rho}_{zz}$ (of the projection on, and the section by, the rectifying plane at O) are zero, infinite or neither, provided that when ρ_{zz} and $\bar{\rho}_{zz}$ are infinite, the polar line at O either coincides with the binormal or lies at infinity; ρ_{zz} and $\bar{\rho}_{zz}$ are zero, if, and only if, R_1 is zero.

If ρ is finite and $\neq 0$ and if the osculator is simple ($\gamma = 1$), then R_1 is finite and $\neq 0$.

The radius of quasi-spherical curvature R_2 at a singular point O is zero, infinite or neither, according as the radii of curvature ρ_{yz} and $\bar{\rho}_{yz}$ (of the projection on, and the section by, the normal plane at O) are zero, infinite or neither, provided that when ρ_{yz} and $\bar{\rho}_{yz}$ are zero, the rectifying line at O coincides either with the binormal or with the tangent; ρ_{yz} and $\bar{\rho}_{yz}$ are infinite, if and only if R_2 is infinite and σ is zero.

If σ is finite and $\neq 0$ and if the point is simple ($\alpha = 1$), then R_2 cannot vanish.

If ρ_{zz} and $\bar{\rho}_{zz}$ are finite and $\neq 0$ $(\alpha=\beta+\gamma)$, we can say more. For by comparing (36), case (b), with (14_1) and (15_1) we see that R_1 is not only finite and $\neq 0$, but is actually equal to ρ_{zz}; also since $\rho=0$ $(\alpha>\beta)$, the polar line coincides with the binormal.

Hence *if the curvature of the projection of the curve on the rectifying plane is finite and different from zero* $(\alpha=\beta+\gamma)$, *then the osculating circle of this projection is a great circle of the corresponding osculating sphere of the original curve.* This sphere will lie on the positive side $(z>0)$ of the osculator.

If ρ_{yz} and $\bar{\rho}_{yz}$ are finite and $\neq 0$ $(\alpha+\beta=\gamma)$, we can say more. For by comparing (39), case (b), with (22) and (22′) we see that R_2 is not only finite and $\neq 0$, but is actually equal to $|\bar{\rho}_{yz}|$; also since $\sigma=0(\gamma>\beta)$, the rectifying line coincides with the binormal.

If the curvature of the section of the tangent surface by the normal plane is finite and different from zero $(\alpha+\beta=\gamma)$, *then the osculating circle of this section is a great circle of the corresponding quasi-osculating sphere of the original curve.* This sphere will lie on the negative side $(z<0)$ of the osculator.

22. If $\alpha=\beta+\gamma$, either all three of the integers α, β, γ are even or just one is even, so that by §4 the point belongs to category 1, 2, 3 or 4. Hence, by §§12 and 15, we see that if the negative branch of the curve lies in one of the octants 5, 6, 7 or 8,

and if $\rho=0$ $(\alpha>\beta)$, then R_1, ρ_{zz}, $\bar{\rho}_{zz}$, ρ_{yz}, and $\bar{\rho}_{yz}$ are all $=0$ or ∞.

and if $\sigma=0$ $(\gamma>\beta)$, then R_2, ρ_{zz}, $\bar{\rho}_{zz}$, ρ_{yz} and $\bar{\rho}_{yz}$ are all $=0$ or ∞.

In other words *if the two branches of the curve at a point O lie on opposite sides of the osculating plane,*

and if the polar line at O coincides with the binormal, then no one of the five variables ρ_{zz}, $\bar{\rho}_{zz}$, ρ_{yz}, $\bar{\rho}_{yz}$, R_1 *can be finite and different from zero.*

If R_1 is finite and $\neq 0$, either $\alpha=\beta+\gamma$ or $\alpha=\beta$. By §8 this can happen only for points of categories 1, 2, 3, 4, 7 or 8.

Hence *if the negative branch of the curve lies in the 5th or 6th octant, no one of the four variables* ρ, ρ_{zz}, $\bar{\rho}_{zz}$, R_1 *can be finite and different from zero.*

and if the rectifying line at O coincides with the binormal, then no one of the five variables ρ_{zz}, $\bar{\rho}_{zz}$, ρ_{yz}, $\bar{\rho}_{yz}$, R_2 *can be finite and different from zero.*

If R_2 is finite and $\neq 0$, either $\alpha+\beta=\gamma$ or $\beta=\gamma$. By §8 this can happen only for points of categories 1, 2, 3, 4, 5 or 8.

Hence *if the negative branch of the curve lies in the 6th or 7th octant, no one of the four variables* σ, ρ_{yz}, $\bar{\rho}_{yz}$, R_2 *can be finite and different from zero.*

An immediate consequence is the remarkable fact that *if the negative branch lies in the 6th octant* $[+1, -1, -1]$, *no one of the eight variables* ρ, σ, ρ_{zz}, $\bar{\rho}_{zz}$, ρ_{yz}, $\bar{\rho}_{yz}$, R_1, R_2 *can be finite and different from zero.* The simplest example of this is a point of type (2, 1, 2).

23. Finally we shall exhibit in the following table some of our results as applied to a number of the simpler types of singular points, together with the type (1, 1, 1) of non-singular points. Here a, b, c, c_1, and h_2 are all finite and $\neq 0$; a, b, c are the coefficients in the equations (5) of §3; h_2 is given by (37_2) of §20; and c_1 is the value of R_1 when it may come under case (d) or case (e) of (36), §19.

The table includes the twenty-seven types for which α, β, γ are $\leqq 3$ and a few others. Wherever three types are grouped together, the first and third are dual and the second is self-dual. All of the types listed belong to distinct species except (1, 1, 1) and (3, 3, 3); the latter is the simplest type of singular point that looks exactly like a non-singular point.

	$\alpha\beta\gamma$	$\alpha, \alpha+\beta, \alpha+\beta+\gamma$	Oct-ant	ρ	τ	σ	ρ_{zz}	$\bar{\rho}_{zz}$	R_1	$\bar{\rho}_{zy}$	ρ_{yz}	$\bar{\rho}_{yz}$	R_2
1	111	123	8	$\frac{a^2}{2b}$	$-\frac{ab}{3c}$	$-\frac{3ac}{2b^2}$	∞	∞	c_1	$\frac{2a^2}{3b}$	0	0	h_2 or ∞
2	112	124	2	$\frac{a^2}{2b}$	∞	0	∞	∞	c_1 or ∞	$\frac{9a^2}{16b}$	$\frac{b^2}{2c}$	$-\frac{b^2}{6c}$	$\frac{b^2}{6c}$
3	121	134	3	∞	$-\frac{ab}{2c}$	∞	∞	∞	∞	∞	0	0	0
4	211	234	4	0	0	$-\frac{16ac}{9b^2}$	$\frac{a^2}{2c}$	$-\frac{a^2}{6c}$	$\frac{a^2}{2c}$	0	0	0	0, h_2 or ∞
5	122	136	5	∞	∞	$-\frac{10ac}{9b^2}$	∞	∞	∞	∞	0	0	h_2 or ∞
6	212	235	6	0	$-\frac{ab}{5c}$	0	∞	∞	∞	0	0	0	0
7	221	245	7	$\frac{a^2}{2b}$	0	∞	∞	∞	c_1	$\frac{9a^2}{10b}$	0	0	0
8	222	246	1	$\frac{a^2}{2b}$	$-\frac{ab}{3c}$	$-\frac{3ac}{2b^2}$	∞	∞	c_1 or ∞	$\frac{2a^2}{3b}$	0	0	0, h_2 or ∞
9	123	136	3	∞	∞	0	∞	∞	∞	∞	$\frac{b^2}{2c}$	$-\frac{2b^2}{5c}$	$\frac{2b^2}{5c}$
10	132	146	2	∞	∞	∞	∞	∞	∞	∞	0	0	0
11	213	236	4	0	∞	0	∞	∞	∞	0	$\frac{b^2}{2c}$	$-\frac{b^2}{16c}$	$\frac{b^2}{16c}$
12	231	256	4	∞	0	∞	∞	∞	∞	∞	0	0	0
13	312	346	2	0	0	0	$\frac{a^2}{2c}$	$-\frac{a^2}{16c}$	$\frac{a^2}{2c}$	0	0	0	0
14	321	356	3	0	0	∞	$\frac{a^2}{2c}$	$-\frac{2a^2}{5c}$	$\frac{a^2}{2c}$	0	0	0	0

	$\alpha\beta\gamma$	$\alpha, \alpha+\beta, \alpha+\beta+\gamma$	Octant	ρ	τ	σ	ρ_{xz}	ρ_{xz}	R_1	ρ_{zz}	ρ_{yz}	ρ_{yz}	R_2
15	113	125	8	$\frac{a^2}{2b}$	∞	0	∞	∞	c_1 or ∞	$\frac{8a^2}{15b}$	∞	∞	∞
16	131	145	8	∞	$-\frac{3ab}{5c}$	∞	∞	∞	∞	∞	0	0	0
17	311	345	8	0	0	$-\frac{15ac}{8b^2}$	0	0	0	0	0	0	0, h_2 or ∞
18	133	147	8	∞	∞	$-\frac{7ac}{8b^2}$	∞	∞	∞	∞	0	0	h_2 or ∞
19	313	347	8	0	$-\frac{ab}{7c}$	0	∞	∞	∞	0	0	0	0
20	331	367	8	$\frac{a^2}{2b}$	0	∞	∞	∞	c_1	$\frac{8a^2}{7b}$	0	0	0
21	223	247	7	$\frac{a^2}{2b}$	∞	0	∞	∞	c_1 or ∞	$\frac{25a^2}{42b}$	0	0	0
22	232	257	6	∞	$-\frac{3ab}{7c}$	∞	∞	∞	∞	∞	0	0	0
23	322	357	5	0	0	$-\frac{42ac}{25b^2}$	∞	∞	∞	0	0	0	0, h_2 or ∞
24	233	258	4	∞	∞	$-\frac{32ac}{25b^2}$	∞	∞	∞	∞	0	0	0, h_2 or ∞
25	323	358	3	0	$-\frac{ab}{4c}$	0	∞	∞	∞	0	0	0	0
26	332	368	2	$\frac{a^2}{2b}$	0	∞	∞	∞	c_1 or ∞	$\frac{25a^2}{32b}$	0	0	0
27	333	369	8	$\frac{a^2}{2b}$	$-\frac{ab}{3c}$	$-\frac{3ac}{2b^2}$	∞	∞	c_1 or ∞	$\frac{2a^2}{3b}$	0	0	0, h_2 or ∞
28	124	137	5	∞	∞	0	∞	∞	∞	∞	∞	∞	∞
29	142	157	5	∞	∞	∞	∞	∞	∞	∞	0	0	0
30	214	237	6	0	∞	0	∞	∞	∞	0	∞	∞	∞
31	241	267	7	∞	0	∞	∞	∞	∞	∞	0	0	0
32	412	457	6	0	0	0	0	0	0	0	0	0	0
33	421	467	7	0	0	∞	0	0	0	0	0	0	0
34	224	248	1	$\frac{a^2}{2b}$	∞	0	∞	∞	c_1 or ∞	$\frac{9a^2}{16b}$	$\frac{b^2}{2c}$	$-\frac{b^2}{6c}$	$\frac{b^2}{6c}$
35	242	268	1	∞	$-\frac{ab}{2c}$	∞	∞	∞	∞	∞	0	0	0
36	422	468	1	0	0	$-\frac{16ac}{9b^2}$	$\frac{a^2}{2c}$	$-\frac{a^2}{6c}$	$\frac{a^2}{2c}$	0	0	0	0, h_2 or ∞

CORNELL UNIVERSITY,
ITHACA, N. Y.

UNIVERSAL QUADRATIC FORMS*

BY

L. E. DICKSON

1. A form is called universal if it represents all integers, and *Null* if it is zero when the variables are integers not all zero. We shall determine all universal Null quadratic forms F in n variables for $n \leqq 4$.

For $n=3$, F is readily reduced to $2^e gaxy+f$, where $f=gby^2+cyz+gdz^2$, ga is odd, a is prime to d, and g to c. Let R be the discriminant of f. Let t be the the largest divisor of a which is prime to g. Then F is universal if and only if R is a quadratic residue of t and one of the following sets of conditions holds: (I) $e=0$; (II) c even, $e=1$, either d is odd, or $d\equiv 2 \pmod 4$ and b is odd; (III) c odd, $e\geqq 1$, bd even. There is a canonical form (§21) which depends only on the Hessian.

For $n=4$ numerous subdivisions arise. There is almost an even chance that a Null form taken at random is universal. The conditions for universality are much milder for $n=4$ than for $n=3$. They are still milder for $n\geqq 5$, the theory for which is under elaboration in a Chicago thesis.

2. **Reduction to normal form.** Let $\xi, \eta, \cdots$ be integral values, not all zero, of the variables $x, y, \cdots$ for which the Null form N vanishes. In view of the homogeneity of N, we may assume that $\xi, \eta, \cdots$ have no common factor >1. It is known that there exists a square matrix M of determinant unity whose elements are all integers, those of the first column being $\xi, \eta, \cdots$. The linear substitution with the matrix M evidently replaces N by a form F in which the coefficient of x^2 is zero.

When there are only two variables, $F=axy+by^2$. The case in which a and b have a common factor $c>1$ is excluded since F then represents only multiples of c.

First, let a have an odd prime factor p. Then F represents no integer of the form $bv+pk$, where v is a quadratic non-residue of p.

Second, let a be even and hence b odd. Then F is never the double of an odd integer.

Hence $a=1$. Replacing x by $x-by$, we get xy.

THEOREM 1. *Every universal binary Null quadratic form is the product of two linear functions of determinant unity and hence is equivalent to* xy.

---

* Presented to the Society, December 31, 1928; received by the editors in September, 1928.

Henceforth let there be n variables, where $n \geqq 3$. The part of F which involves x may be written as Axy', where y' is a linear function of $y, z, \cdots$, the greatest common divisor of whose coefficients is unity. As noted above there exists a square matrix of determinant unity whose elements are all integers, those of the first row being the coefficients of y'. Let $z', w', \cdots$ be the linear functions of $y, z, \cdots$ whose coefficients are the elements of the second, third, $\cdots$ rows of that matrix. The resulting linear substitution replaces F by an equivalent form f. After dropping the accents on y', $z', \cdots$, we have

$$f = Axy + \phi(y,z,w, \cdots). \tag{1}$$

If $n=3$, this is of the form (2) with $\psi = z^2$. If $n>3$, the sum of the terms of ϕ which are linear in y is cyz', where z' is a linear function of $z, w, \cdots$, the g. c. d. of whose coefficients is 1. There exist further linear functions $w', \cdots$ such that the determinant of the coefficients in $z', w', \cdots$ is 1. Hence f is equivalent to

$$h = Axy + By^2 + cyz + \Delta\psi(z,w, \cdots), \tag{2}$$

in which the g. c. d. of the coefficients of ψ is 1.

Let $A = 2^e\alpha$, where α is odd. Let g be the g. c. d. of $\alpha = ga$ and $\Delta = gd$. Then c is prime to g. For, if a prime p divides c and g, p is odd and $h \equiv By^2 \pmod{p}$, whence h has at most $\frac{1}{2}(p+1)$ values modulo p and is not universal.

If we replace z by $z+ty$ in h and note that $\Delta \equiv 0 \pmod{g}$, we get a form (1) in which the coefficient of y^2 is $\equiv B+ct \pmod{g}$. This is divisible by g when t is suitably chosen. Hence we may take $B = gb$ in (2).

THEOREM 2. *Every universal Null quadratic form in three or more variables is equivalent to a form*

$$F = 2^e gaxy + gby^2 + cyz + gd\psi(z,w, \cdots), \tag{3}$$

where g and a are odd, a is prime to d, c is prime to g, and the g. c. d. of the coefficients of ψ is unity.

PART I. CASE OF THREE VARIABLES

3. Here $F = Px + f$, $P = 2^e gay$, $f = gby^2 + cyz + gdz^2$. Let G be any given integer. Our method is briefly as follows. Under specified conditions on the coefficients of F, we shall show how to select an odd prime π, not dividing gad, such that $f \equiv G \pmod{P}$ has a solution $z = Z$ when $y = \pi$. Thus $f = G + PQ$, where Q is an integer. Hence $F = G$ when $x = -Q$, $y = \pi$, $z = Z$.

If $f \equiv G$ is solvable for the separate moduli 2^e, ga, y, it is solvable modulo P. For modulus y, the condition is

$$(gdz)^2 \equiv gdG \pmod{y}. \tag{4}$$

We shall satisfy this condition in §6 by choice of $y=\pi$.

Consider modulus ga. A prime factor of it either divides both g and a or just one of them. Hence we may write

$$(5)\qquad \begin{cases} g=qr,\ a=st,\ q \text{ and } s \text{ have the same distinct prime factors,} \\ r \text{ and } t \text{ are prime to each other and to both } q \text{ and } s. \end{cases}$$

If $f\equiv G$ is solvable for the separate moduli qs, r, t, which are relatively prime in pairs, it is solvable modulo ga, their product. By (5) and Theorem 2,

$$(6)\qquad d \text{ is prime to } st, \quad c \text{ is prime to } q, r, \text{ and } s.$$

Since $f\equiv cyz \pmod g$, $f\equiv G \pmod g$ is solvable when $y=\pi$ by (6) and the fact that π does not divide g. This disposes of modulus r.

4. Consider $f\equiv G \pmod{qs}$. Since q is a factor of g, we saw that $f\equiv cyz\equiv G \pmod q$ has a solution z_1 when $y=\pi$, whence $cyz_1=G+Mq$. Its general solution is $z=z_1+\zeta q$, where ζ is arbitrary. Insert the value of z into $f\equiv G \pmod{qs}$, cancel G, and delete the common factor q. We get

$$rby^2+cy\zeta+M+rd(z_1+\zeta q)^2\equiv 0 \pmod s.$$

If s is a product of powers p^n of distinct primes, it suffices to prove that the like congruence is solvable for each modulus p^n. Let p^m be the highest power of p which divides q, whence $m\geqq 1$. The congruence is of the form

$$(7)\qquad S\zeta+p^m\phi(\zeta)\equiv k \pmod{p^n},$$

where $S=cy$ is not divisible by p by (6), and k depends upon y, but not on ζ. If $m\geqq n$, this is $S\zeta\equiv k$ and is solvable. Next, let $n>m$. As before, (7) has a solution ζ' modulo p^m, whence

$$\zeta=\zeta'+Zp^m, \quad S\zeta'=k+Rp^m.$$

Cancellation of k from (7) and division by p^m gives

$$R+SZ+\phi(\zeta'+Zp^m)\equiv 0, \text{ or } SZ+p^mP(Z)\equiv k' \pmod{p^{n-m}}$$

where $k'=\phi(\zeta')-R$ is independent of Z. If $n-m\leqq m$, this is $SZ\equiv k'$ and is solvable. If $n-m>m$, we repeat the process and reduce the problem to a congruence modulo p^{n-2m}.

To proceed by induction on μ, suppose the problem has been reduced to

$$(8)\qquad Su+p^m\phi(u)\equiv k \pmod{p^{n-\mu m}}, \quad n>\mu m.$$

If $n-\mu m\leqq m$, then p^m is a multiple of the modulus and (8) is solvable. Next, let $n-\mu m>m$. Evidently (8) has a solution u' modulo p^m, whence

$$u=u'+vp^m, \quad Su'=k+p^mQ.$$

Cancellation of k from (8) and division by p^m gives

$$Sv + Q + \phi(u' + vp^m) \equiv 0, \quad \text{or} \quad Sv + p^m P(v) \equiv k' \pmod{p^{n-\mu m-m}},$$

where k' is free of v. Since this is of type (8) with μ replaced by $\mu+1$, the induction is complete. Ultimately we reach a congruence (8) with $n-\mu m \leqq m$, which is therefore solvable. This proves

LEMMA 1. *For every integer G, and for $y=\pi$, $F\equiv G$ is always solvable modulis qs and r.*

5. Let t be a product of powers p^n of distinct primes. Multiplication of $f\equiv G \pmod{p^n}$ by $4gd$, which is prime to t by (5) and (6), gives the equivalent congruence

$$Z^2 - Ry^2 \equiv k \pmod{p^n}, \tag{9}$$

where

$$Z = 2gdz + cy, \quad R = c^2 - 4g^2db, \quad k = 4gdG.$$

If $R\equiv 0 \pmod{p}$, $4gdF \equiv Z^2$, whence F has only $\frac{1}{2}(p+1)$ values modulo p and is not universal.

Let R be a quadratic non-residue of p, and take $G\equiv 0$, whence $k\equiv 0 \pmod{p}$. Then $y\equiv 0$, $Z\equiv 0$, $z\equiv 0 \pmod{p}$. Thus F is divisible by p^2 and is not universal.

Hence $v^2\equiv R \pmod{p^n}$ is solvable when $n=1$. To proceed by induction from $n=m$ to $n=m+1$, let $w^2 = R+Sp^m$. Then w is not divisible by p, and $2wT+S\equiv 0 \pmod{p}$ has a solution T. Hence $(w+Tp^m)^2 \equiv R \pmod{p^{m+1}}$.

Determine δ by $v\delta\equiv 1 \pmod{p^n}$. Multiply (9) by δ^2 and write $u=\delta Z$, $K=\delta^2 k$. We get

$$u^2 - y^2 \equiv K \pmod{p^n}. \tag{10}$$

Modulo p, this has a solution with y prime to p unless

$$p = 3, \quad K \equiv 1 \pmod{3}. \tag{11}$$

To prove this fact, let α be any integer not divisible by p and determine β by $\alpha\beta\equiv 1 \pmod{p}$. Then $2y\equiv \alpha - K\beta \pmod{p}$ determines an integer y not divisible by p if $\alpha^2 \not\equiv K \pmod{p}$. Since at most three residues α are excluded, we can find a suitable α if $p>3$. In case $p=3$ and $K\equiv 0$ or $2 \pmod{3}$, only $\alpha\equiv 0$ is excluded. Take $u=y-\alpha$. Then $y+u\equiv -K\beta$, and (10) holds.

To show that (10) has a solution with y prime to p, we proceed by induction from $n=m$ to $n=m+1$. Hence let

$$u^2 - V^2 = K + Sp^m, \quad V \text{ prime to } p.$$

Then $2Y\eta \equiv S \pmod{p}$ has a solution η, and (10) holds modulo p^{m+1} for $y = Y + \eta p^m$. Except in case (11), there is therefore an integer Y prime to p such that, when $y \equiv Y \pmod{p^n}$, (10) has a solution u, and hence $f \equiv G \pmod{p^n}$ has a solution z.

In case $G \equiv 0$, whence $K \equiv 0 \pmod{p}$, we shall need the fact that (10) has a solution in which y has any assigned value v not divisible by p. If $n = 1$, we may take $u = v$. To proceed by induction, let

$$U^2 - v^2 = K + p^m Q.$$

Then (10) holds modulo p^{m+1} when $u = U + Sp^m$, $y = v$, if $2SU + Q \equiv 0 \pmod{p}$, which is satisfied by choice of S.

6. We are now in a position to prove

LEMMA 2. *If R is a quadratic residue of t, to each G corresponds an odd prime π not dividing agd such that, when $y = \pi$, $f \equiv G \pmod{ty}$ is solvable.*

Write $t = \tau T$, $\tau = p_1^{n_1} \cdots p_k^{n_k}$, where no one of the distinct primes p_i divides G, while each prime factor of T divides G. Except in case (11), we saw that there is an integer Y_i not divisible by p_i such that, when $y \equiv Y_i \pmod{p_i^{n_i}}$, there exists a solution z_i of $f \equiv G \pmod{p_i^{n_i}}$. But there are integers Y and z satisfying

$$Y \equiv Y_1, \quad z \equiv z_1 \pmod{p_1^{n_1}}, \quad \cdots, \quad Y \equiv Y_k, \quad z \equiv z_k \pmod{p_k^{n_k}}.$$

Hence Y is prime to τ, and there is a solution z of $f \equiv G \pmod{\tau}$ when $y = Y$.

Write D for gdG. Since gd is prime to t by (5) and (6), τ is prime to D. The divisor τ of a is odd. Let $\pi_1, \cdots, \pi_h$ be the distinct odd primes which occur in D with odd exponents. The system of congruences

$$\pi \equiv Y \pmod{\tau}, \quad \pi \equiv 1 \pmod{8}, \quad \pi \equiv 1 \pmod{\pi_i} \quad (i = 1, \cdots, h)$$

has a solution $\pi \equiv V \pmod{M}$, where V is prime to $M = 8\tau\pi_1 \cdots \pi_h$. There are infinitely many primes of the form $V + Mw$. Let π be one of them which does not divide $2\,agd$. We shall prove that Lemma 2 holds for this π. For Jacobi's symbols,

$$(\pi_i/\pi) = (\pi/\pi_i) = 1, \quad (2/\pi) = 1, \quad (D/\pi) = 1.$$

Hence by (4), $f \equiv G \pmod{y}$ has a solution z when $y = \pi$. We saw that $f \equiv G \pmod{\tau}$ has a solution z when $y \equiv Y \equiv \pi \pmod{\tau}$. It remains only to prove that $f \equiv G \pmod{T}$ has a solution z when $y = \pi$. Let p^n be a highest power of a prime dividing T. Thus p divides G. Since π is not a divisor of a and hence not of T, $\pi \neq p$. At the end of §5, we saw that, when y has any assigned value not divisible by p, and hence when $y = \pi$, $f \equiv G \pmod{p^n}$ has a solution z. The same is true modulo T.

Combining Lemmas 1 and 2, we have, except for case (11),

THEOREM 3. *Let t denote the largest divisor of a which is prime to g. If F is universal, then*

$$R=c^2-4g^2db \text{ is a quadratic residue of } t. \tag{12}$$

In case (12), *to every G corresponds an odd integer π such that, when $y=\pi$, $f\equiv G \pmod{agy}$ has a solution z.*

7. It remains to prove Theorem 3 for the special case (11). Then $\delta^2\equiv 1$, $R\equiv 1$, $k\equiv 1 \pmod 3$, and (9) requires $y\equiv 0$. Take Y prime to 3, and $3Y\equiv Y_i \pmod{p_i^{n_i}}$ for each prime factor $p_i\neq 3$ of τ. Now $k+9RY^2$ is a quadratic residue of 3 and hence of 3^{n+1}. For $y\equiv 3Y$, $f\equiv G$ therefore has a solution z modulo 3^{n+1} and hence modulo 3τ. Define π as in §6. For $y\equiv 3\pi$, $f\equiv G$ has a solution z modulo 3τ and modulo π, and hence modulo $3\tau\pi=y\tau$.

By the first remark in §3 and Theorem 3, we have

THEOREM 4. *If $e=0$, F is universal in case* (12).

8. Consider the classic case of forms F in which the coefficients of products of different variables are all even, whence $e\geqq 1$ and c is even. We shall prove

THEOREM 5. *When $e\geqq 1$ and c is even, F is universal if and only if* (12) *holds and*

$$e=1;\ \textit{either } d \textit{ is odd, or } d\equiv 2 \pmod 4 \textit{ and } b \textit{ is odd.} \tag{13}$$

First, let F be universal and employ the notations $A=ga$, $B=gb$, $c=2C$, $D=gd$. Then

$$F=2^eAxy+f, \quad f=By^2+2Cyz+Dz^2, \quad A \text{ odd}.$$

Since F shall represent odd integers, B and D are not both even. If B is even and D odd, we replace z by $z+y$ in F and obtain a like form with $B'=B+2C+D$, which is odd. Hence we may take B odd.

First, let $e\geqq 3$. Since F represents a complete set of residues modulo 8, the same is true of

$$BF\equiv y^2+2BCyz+BDz^2\equiv Y^2+hz^2 \pmod 8,$$

where $Y=y+BCz$. If h is even, Y^2+hz^2 has at most six values modulo 8. Hence h is odd and Y^2+hz^2 has at most seven residues $0, 1, 4; h, h+1, h+4; 5$.

Second, let $e=2$. If $A\equiv 3 \pmod 4$, we change the signs of y and z in F and obtain an equivalent form having $A'=-A$. Hence let $A\equiv 1 \pmod 4$.

Then $F \equiv 4xy+f \pmod{16}$. Replacing x by $x+ky+sz$, we obtain a like form having $B'=B+4k$, $C'=C+2s$. Hence we may take $B=\pm 1$, $C=0$ or 1.

The case $B=C=1$. The residues of F modulo 4 are 0, 1, D, $D+3$. Hence $D=4k+3$. Consider odd values of F. Then $y+z$ is odd and

$$f = (y+z)^2 + (D-1)z^2, \quad F \equiv 4x(z+1) + 1 + (4k+2)z^2 \pmod 8.$$

For z even, $F \equiv 4x+1 \equiv 1$ or 5; for z odd, $F \equiv 4k+3 \pmod 8$. Hence F represents only three of the four odd classes modulo 8.

The case $B=1$, $C=0$. Then $F \equiv 0$, 1, D, $D+1 \pmod 4$, whence $D=4k+2$. Then

$$F \equiv 4xy + y^2 + (4k+2)z^2 \pmod{16}.$$

When F is even, $y=2Y$ and $F=2\phi$, $\phi = 4xY+2Y^2+mz^2$, $m=2k+1$. Since F represents all even residues modulo 16, ϕ represents all residues modulo 8. But if ϕ is odd, z is odd and $\phi \equiv m$, $m+2$, or $m+6 \pmod 8$. Thus $\phi \not\equiv m+4 \pmod 8$.

The case $B=-1$. In the universal form $-F$ we change the signs of y and D and obtain F with $B=1$, which was treated in the preceding cases.

This proves that $e=1$. We return to the notations in §3. By hypothesis, c is even and g is odd. Let d be even. Since F is not always even, b is odd. If $d \equiv 0 \pmod 4$, $F \equiv 2 \pmod 4$ is impossible. For that requires that y be even and then $F \equiv 0 \pmod 4$. Hence (13) are necessary conditions that F be universal.

We readily show that (12) and (13) are sufficient conditions. If d is odd, and y is any chosen odd integer, $F \equiv by^2+dz^2 \equiv G \pmod 2$ has a solution z when G is arbitrary. Thus F is universal by Theorem 3. Next, let $d=2D$, where D and b are odd. If y is odd, then F is odd and $F \equiv G \pmod 2$ has a solution z when G is odd. Next, let G be even. Take $y=2Y$, where Y is odd. Then $F = 4gaxY+4gbY^2+2cYz+gdz^2 \equiv 2z^2 \pmod 4$, whence $F \equiv G \pmod 4$ has a solution z. This F is derived from (3) by replacing e by 2, b by $4b$, c by $2c$, y by Y, and has the same g, a, d. The conditions in Theorem 2 still hold, while R is multiplied by 4. We may therefore apply Theorem 3 with $Y=\pi$, $e=2$.

9. In view of Theorems 4 and 5, it remains only to treat the case $e \geqq 1$, c odd.

(I) Let d be even. Assign an odd value to y and write k for the odd integer cy. Then $F \equiv gby^2+\phi \pmod{2^e}$, where $\phi = kz+gdz^2$. Then ϕ ranges with z over a complete set of residues modulo 2^e. For, $\phi \equiv kZ+gdZ^2$ implies

$$(z-Z)[k+gd(z+Z)] \equiv 0 \pmod{2^e}.$$

Since the second factor is odd, $z \equiv Z \pmod{2^e}$. Hence if G is arbitrary, $F \equiv G \pmod{2^e}$ has a solution z. Hence F is universal if (12) holds.

(II) Let d be odd. If b is odd, $F \equiv y + yz + z \pmod 2$ and F is even only when y and z are both even, whence $F \not\equiv 2 \pmod 4$. Hence for a universal F, b is even.

Determine ω so that $gd\omega \equiv 1$ modulo 2^{e+3}; then $F \equiv G$ is solvable if and only if $\omega F \equiv \omega G$ is solvable. It therefore suffices to study

$$H = 2^e Axy + 2By^2 + Cyz + z^2 \qquad (e \geqq 1,\ C \text{ odd}). \tag{14}$$

(i) Assign a fixed odd value to y. The values of H modulo 2^e are the sums of $2By^2$ and the values of $\phi = z^2 + tz$, where $t = Cy$ is odd. Evidently ϕ is always even. Consider

$$\phi \equiv Z^2 + tZ, \quad (z - Z)(z + Z + t) \equiv 0 \pmod{2^e}.$$

If the first factor is even, the second is odd and $z \equiv Z \pmod{2^e}$. Hence if z ranges over the 2^{e-1} even integers

$$0, 2, 4, \cdots, 2^e - 2, \tag{15}$$

ϕ takes 2^{e-1} even values incongruent modulo 2^e, which are therefore congruent to the numbers (15) rearranged. Since this result is not changed if we add to them the constant $2By^2$, we conclude that $H \equiv k \pmod{2^e}$ has a solution when k is any even integer. Hence Theorem 3 applies when G is any even integer.

(ii) Let $y = 2y_1$, where y_1 is odd, but undetermined. Then

$$H \equiv Z^2 + ry_1^2 \pmod{2^{e+1}}, \quad Z = z + Cy_1, \quad r = 8B - C^2.$$

In view of (i), we need consider only odd values of H. Then $Z = 2\zeta$ and $H \equiv 3 \pmod 4$. If k is any integer $\equiv 3 \pmod 4$,

$$4\zeta^2 + ry_1^2 \equiv k \pmod{2^{e+1}}$$

has a solution with y_1 odd. This is evidently true modulo 8. To proceed by induction, let

$$4a^2 + rb^2 = k + 2^m Q, \quad b \text{ odd}, \quad m \geqq 3.$$

Then

$$4a^2 + r(b + 2^{m-1}w)^2 \equiv k + 2^m(Q + rbw) \equiv k \pmod{2^{m+1}},$$

by choice of w modulo 2. The induction is complete. To each k corresponds an odd G for which $F \equiv G \pmod{2^{e+1}}$ has a solution with y_1 odd.

Since y_1 was not preassigned, but was suitably determined, we must modify our determination of π in §6. Since $D = gdG$ is now odd, we omit

$\pi \equiv 1 \pmod 8$ from the system of congruences for π and replace it by $\pi \equiv y_1 \pmod{2^{e+1}}$.

(iii) Let $y = 4y_2$, where y_2 is a fixed odd integer. Then

$$H \equiv Z^2 + 4ry_2^2 \pmod{2^{e+2}}, \qquad Z = z + 2Cy_2, \qquad r = 8B - C^2.$$

In view of (i), we need consider only odd values of H, whence Z is odd and $H \equiv 5 \pmod 8$. If k is any integer $\equiv 5 \pmod 8$, then $H \equiv k \pmod{2^{e+2}}$ has an odd solution Z. This is evident for modulus 8. To proceed by induction, let

$$a^2 + 4ry_2^2 = k + 2^m Q, \quad a \text{ odd}, \quad m \geqq 3.$$

Then

$$(a + 2^{m-1}x)^2 + 4ry_2^2 \equiv k + 2^m(Q + ax) \equiv k \pmod{2^{m+1}},$$

by choice of x modulo 2. Our F is derived from (3) by replacing e by $e+2$, y by y_2, b by $16b$, and c by $4c$, and has the same g, a, d. The conditions in Theorem 2 hold also here, while R is multiplied by 16.

(iv) Let $y = 8y_3$, where y_3 is a fixed odd integer. Then

$$H \equiv Z^2 + 16ry_3^2 \pmod{2^{e+3}}, \quad Z = z + 4Cy_3.$$

Take Z odd. Then $H \equiv 1 \pmod 8$. If k is any integer $\equiv 1 \pmod 8$, then $H \equiv k \pmod{2^{e+3}}$ has an odd solution Z. This is proved by induction as in (iii).

Since every integer k falls under one of our four cases, we conclude from Theorem 3 that F is universal if (12) holds. This completes the proof of

THEOREM 6. *When $e \geqq 1$ and c is odd, F is universal if and only if* (12) *holds and bd is even.*

10. Another proof of Theorem 6 reveals a property to be utilized in the more complicated case of four variables. When (12) holds, F is universal if and only if $F \equiv G \pmod{2^n}$ is solvable when $n \leqq 2e+2$, irrespective of the evenness or oddness of y.

When G is even, we retain the proof in (i) of §9. Next, let G be odd. Now H is the product of F by ω modulo 2^{2e+2} instead of 2^{e+3}. For n arbitrary and k odd, $H \equiv k \pmod{2^n}$ has a solution with $x = 0$, y even. For proof, put $y = 2Y$. Then

$$H = Z^2 + rY^2, \qquad Z = z + CY, \qquad r = 8B - C^2.$$

Modulo 8, $H \equiv Z^2 - Y^2$ has the values 0, 1, 3, 4, 5, 7, whence $H \equiv \text{odd} \pmod 8$ is solvable. To proceed by induction, let

$$\xi^2 + r\eta^2 = k + 2^m Q, \quad m \geqq 3.$$

Since ζ and η are not both even,

$$(\zeta + 2^{m-1}u)^2 + r(\eta + 2^{m-1}v)^2 \equiv k + 2^m(Q + \zeta u + r\eta v) \equiv k \pmod{2^{m+1}},$$

by choice of u, v modulo 2.

Now take $n=2e+2$. If in a solution of $H\equiv k \pmod{2^n}$, Y is divisible by $h=2^{e+1}$, then $Z^2\equiv k$, whence $H\equiv k \pmod{2^n}$ has a solution in which Y is an arbitrary multiple of h, and hence a solution with $Y=h$. Next, let there be no solution having Y divisible by h. Then in every solution, Y is the product of 2^s by an odd integer where $s\leqq e$. In both cases there is a solution with $Y=2^\sigma\eta$, where η is odd and $\sigma\leqq e+1$. Then $y=2^l\eta$, $l\leqq e+2$. Since $e+l\leqq n$, we see that $H\equiv k \pmod{2^{e+l}}$ has a solution with $x=0$, $y=2^l\eta$. Insertion of this y into (3) gives a form F_1 which is derived from F by replacing e by $e+l$, y by η, b by $2^{2l}b$, and c by 2^lc. Since F_1 has the same g, a, d as F, F_1 satisfies the conditions in Theorem 2. The R of F_1 is the product of R in (12) by 2^{2l}. For G odd, we proved that $F=F_1\equiv G \pmod{2^{e+l}}$ has a solution with η odd. Thus F_1 and therefore F is universal if (12) holds.

PART II. THE CASE OF FOUR VARIABLES

11. In (3), let

$$\psi = hz^2 + jzw + lw^2, \quad 1 = \text{g.c.d. of } h,j,l. \tag{16}$$

We may assume that *h is relatively prime to any given odd integer m.* For, the replacement of w by $w+tz$ alters only h and j. Then $h'=h+jt+lt^2$. We can choose t so that h' is divisible by no one of the distinct prime factors $p_1, \cdots, p_k$ of m. In fact, since h, j, l are not all divisible by p_i, there are at most two incongruent roots t of $h'\equiv 0 \pmod{p_i}$. Since $p_i>2$, there is a value v_i of t such that h' is not divisible by p_i. There exists an integer v such that

$$v \equiv v_1 \pmod{p_1}, \cdots, v \equiv v_k \pmod{p_k}.$$

Hence when $t=v$, h' is divisible by no one of $p_1, \cdots, p_k$.

12. Let F have the properties in Theorem 2. In (16) we may take h prime to ga by §11.

LEMMA 3. *If each of the congruences*

$$F \equiv G \pmod{2^s}, \qquad F \equiv G \pmod{ga}$$

has a solution x, y, z, w such that y has a fixed value 1 or an odd prime dividing no one of g, a, d, h, $N=j^2-4hl$, then $F=G$ is solvable.

We first prove that $F\equiv G \pmod{y}$ is solvable. Proof is needed only when

y is the specified odd prime. Determine m by $gdm \equiv 1 \pmod{y}$. Multiplication of $F \equiv G$ by $4mh$ yields the equivalent congruence

$$4h\psi \equiv 4mhG \quad \text{or} \quad Z^2 - Nw^2 \equiv 4mhG \pmod{y},$$

where $Z = 2hz + jw$. There is a solution Z, w since N is not divisible by the odd prime y. Since h is not divisible by y, Z determines z.

Hence $F \equiv G \pmod{2^e gay}$ is solvable. As at the beginning of §3, the equation $F = G$ is solvable.

The proof of Lemma 1 applies also here if we take $w = 0$ and multiply the coefficient of z^2 in §4 by h.

Since t is prime to g, d, and h, while t divides a, multiplication of $F \equiv G \pmod{t}$ by $4gdh$ yields the equivalent congruence

$$(17) \qquad 4g^2dhby^2 + 4gdhcyz + g^2d^2[(2hz + jw)^2 - Nw^2] \equiv 4gdhG \pmod{t}.$$

Let t be a product of powers p^n of distinct primes.

13. **Case N not divisible by p.** The product of (17) by N is

$$(18) \qquad Nu^2 - v^2 + Jy^2 \equiv k \pmod{p^n},$$

where

$$u = gd(2hz + jw) + cy, \qquad v = Ngdw + cjy, \qquad k = 4NgdhG,$$

$$(19) \qquad R = c^2 - 4g^2dhb, \qquad J = c^2j^2 - NR.$$

(I) $J \not\equiv 0 \pmod{p}$. Since $Nu^2 - v^2 \equiv k - J \pmod{p}$ is solvable, (18) has a solution modulo p with $y \equiv 1 \pmod{p}$. Write $Nu^2 - v^2 + J = k + pQ$. Determine c_1 so that $Q + 2Jc_1 \equiv 0 \pmod{p}$. Then (18) holds modulo p^2 with $y \equiv 1 + c_1p \pmod{p^2}$. Hence

$$Nu^2 - v^2 + J(1 + c_1p)^2 = k + p^2T.$$

Determine c_2 so that $T + 2Jc_2 \equiv 0 \pmod{p}$. Then (18) holds modulo p^3 with $y \equiv 1 + c_1p + c_2p^2 \pmod{p^3}$. To proceed by induction from $n = m$ to $n = m + 1$, let (18) hold modulo p^m when $y \equiv Y \pmod{p^m}$, where $Y = 1 + c_1p + \cdots + c_{m-1}p^{m-1}$. Write

$$Nu^2 - v^2 + JY^2 = k + p^mS.$$

Determine c_m so that $S + 2Jc_m \equiv 0 \pmod{p}$. Then (18) holds modulo p^{m+1} with $y \equiv Y + c_mp^m \pmod{p^{m+1}}$. The induction is therefore complete and shows that (18) has solutions with $y \equiv \eta \pmod{p^n}$, $\eta = 1 + c_1p + \cdots + c_{n-1}p^{n-1}$, with each c_i determined modulo p. There exist infinitely many primes y of the form $\eta + xp^n$.

(II) N a quadratic residue of p. Thus $N \equiv T^2 \pmod{p}$. Write U for Tu, K for $k-J$. Take $y=1$. Then (18) holds modulo p if $U^2-v^2 \equiv K \pmod{p}$. This has solutions. In case $K \equiv 0$, take $U \equiv v \equiv 1$. Hence $Nu^2-v^2 \equiv K$ always has solutions u, v, not both divisible by p. To proceed by induction from $n=m$ to $n=m+1$, let

$$Nu^2 - v^2 \equiv K \pmod{p^m}$$

have solutions u, v, not both divisible by p. Then

$$Nu^2 - v^2 = K + p^m Q,$$

$$N(u + p^m\xi)^2 - (v + p^m\eta)^2 \equiv K + p^m L \pmod{p^{m+1}},$$

where $L=Q+2Nu\xi-2v\eta \equiv 0 \pmod{p}$ has solutions ξ, η. Since the induction is complete, (18) has solutions with $y=1$.

(III) If N is a quadratic non-residue of p and $J \equiv 0 \pmod{p}$, F is not universal. Consider (18) for $k=pK$ and write $J=Tp$. Then $Nu^2-v^2 \equiv 0$, $u \equiv v \equiv 0 \pmod{p}$. By the origin of (18),

$$4gdhN(F - G) = Nu^2 - v^2 + Jy^2 - k.$$

The second member is $\equiv pM \pmod{p^2}$, where $M=Ty^2-K$. We can choose K so that M is not divisible by p for any y. To each K corresponds a single G by the value of k below (18). Hence F is never congruent modulo p^2 to certain multiples G of p.

14. **Case** $N \equiv 0 \pmod{p}$. Write $N=p\epsilon$.

(I) Let $jc \not\equiv 0 \pmod{p}$. In (17) we may solve $2hz \equiv -jw \pmod{p^n}$ for z, since h is prime to ga and hence to p. Take $y=1$ and write

$$\mu = 2gdcj, \qquad \nu = g^2d^2\epsilon, \qquad k = 4g^2dhb - 4gdhG.$$

Then (17) is equivalent to

$$\mu w + \nu p w^2 \equiv k \pmod{p^n}, \qquad \mu \not\equiv 0 \pmod{p}. \tag{20}$$

This has a solution w' modulo p, and $w=w'+\omega p$, $\mu w'=k+\gamma p$. Then (20) is equivalent to

$$\mu\omega + \gamma + \nu(w' + \omega p)^2 \equiv 0 \quad \text{or} \quad \mu\omega + pf(\omega) \equiv K \pmod{p^{n-1}}.$$

Suppose we have similarly reduced the solution of (20) to

$$\mu u + pf(u) \equiv K \pmod{p^{n-m}}. \tag{21}$$

This has a solution u' modulo p, and

$$u = u' + vp, \quad \mu u' = K + \delta p.$$

Then (21) is equivalent to

$$\mu v + \delta + f(u' + vp) \equiv 0 \quad \text{or} \quad \mu v + pP(v) \equiv K' \pmod{p^{n-m-1}}.$$

This is of type (21) with m replaced by $m+1$. Hence the induction from m to $m+1$ is complete, and (20) is solvable.

(II) Let $j \equiv 0 \pmod p$. Then (17) gives

$$(22) \qquad Z^2 - Ry^2 \equiv k \pmod p, \qquad Z = 2gdhz + cy, \qquad k = 4gdhG.$$

Since this is of type (9), R is not divisible by p. Next, if R is a quadratic non-residue of p, and if $G \equiv 0 \pmod p$, then $y \equiv 0$, $Z \equiv 0$, $z \equiv 0 \pmod p$, $F \equiv gd\psi \pmod{p^2}$. Since $N = p\epsilon$ and $\zeta = 2hz + jw$ is divisible by p,

$$4hF = gd(\zeta^2 - Nw^2) \equiv \tau p w^2 \pmod{p^2}, \quad \tau = -gd\epsilon.$$

Hence F represents only those multiples mp of p for which $4hm \equiv \tau w^2 \pmod p$. Hence m has at most $\frac{1}{2}(p+1)$ values modulo p. Thus F is not universal.

Hence R must be a quadratic residue of p. We take $w = 0$. The discussion in §§5, 7 applies here. There are infinitely many primes y having specified residues with respect to odd moduli p^n.

(III) Let $c \equiv 0$, $j \not\equiv 0 \pmod p$. In (17) write Z for $gd(2hz + jw)$. We get the congruence (22). As in (II), R must be a quadratic residue of p. By §5 with $n = 1$, (22) then has a solution with y prime to p except in case (11).

There is a solution with $w = 0$, y prime to p, of

$$(23) \qquad F \equiv G\colon\ gby^2 + cyz + gdhz^2 \equiv G \pmod{p^n}.$$

To proceed by induction from $n = m$ to $n = m+1$, let

$$gbY^2 + cYZ + gdhZ^2 = G + kp^m, \quad Y \text{ prime to } p.$$

Then (23) holds modulo p^{m+1} for $y = Y + \eta p^m$, $z = Z$, if

$$k + 2gbY\eta + c\eta Z \equiv 0 \pmod p.$$

This has a solution η since $c \equiv 0$, $R \not\equiv 0$, whence $gb \not\equiv 0$ by (19).

15. This completes the proof of

THEOREM 7. *For the form F defined by* (3) *and* (16), *let g and a be odd, a prime to d, c prime to g, and h prime to ga. Employ the abbreviations*

$$(24) \qquad N = j^2 - 4hl, \quad R = c^2 - 4g^2dhb, \quad J = c^2j^2 - NR.$$

We may assign to y a fixed value which is 1 *or an odd prime such that $F \equiv G \pmod{gay}$ is solvable for every G except in the following cases*:

$$(25) \qquad \begin{aligned} &J \equiv 0 \pmod p,\ (N/p) = -1;\ N \equiv cj \equiv 0 \pmod p \text{ and either} \\ &R \equiv 0 \pmod p \text{ or } (R/p) = -1, \end{aligned}$$

where p is any prime dividing a but not g. In these cases F is never universal.

By the first remark in §3, we have

THEOREM 8. *If $e=0$, F is universal except in cases* (25).

16. Assume that the coefficients of products of different variables in F are all even, as in the classic theory. Hence $e\geqq 1$, c and dj are even. First, let $e=1$.

If d is odd, j is even and h and l are not both even by (16). Then $F=by+hz+lw\equiv G \pmod 2$ is solvable when y has an assigned odd value and G is arbitrary. Then F is universal except in cases (25).

Next, let $d=2D$. Then b must be odd. For G odd, $F\equiv y^2\equiv G \pmod 2$ holds if y is odd, whence §15 applies. Finally, let $G=2\gamma$. Then must $y=2Y$ and $F=2f$, where

$$f = 2gaxY + 2gbY^2 + cYz + gD(hz^2 + jzw + lw^2).$$

Since $f=\gamma$ shall be solvable for every γ, D must be odd. The function in parenthesis can be made congruent to either 0 or 1 modulo 2. Whatever be Y or γ, $f\equiv\gamma \pmod 2$ is therefore solvable. Since f is derived from F by replacing b and d by $2b$ and $\frac{1}{2}d$, the initial conditions in Theorem 7 are satisfied by f, and N, R, J are unaltered. Hence f and F are universal except in cases (25).

THEOREM 9. *Let $e=1$. If d is odd, let c and j be even; then F is universal except in cases* (25). *If $d=2D$, let c be even. Necessary and sufficient conditions that F be universal are that b and D be both odd except in cases* (25).

17. Let $e\geqq 2$, d odd. Then $c=2C$, $j=2s$. The products of h, s, l by the odd interger gd will be designated H, S, L. Write β for gb. Then

$$F = 2^e gaxy + \beta y^2 + 2Cyz + Hz^2 + 2Szw + Lw^2, \tag{26}$$

where H and L are not both even. Here let L be odd. Write

$$\lambda = LH - S^2, \qquad W = Lw + Sz.$$

Since our moduli are powers of 2, we may take W and z as new variables in place of w, z. We get

$$LF = L(2^e gaxy + \beta y^2 + 2Cyz) + \lambda z^2 + W^2. \tag{27}$$

Here let also λ be odd. Write

$$A = \lambda Lga, \qquad M = \lambda L\beta - L^2C^2, \qquad Z = \lambda z + LCy, \qquad F_1 = \lambda LF.$$

Then

$$F_1 = 2^e Axy + My^2 + Z^2 + \lambda W^2 \qquad (A, \lambda \text{ odd}). \tag{28}$$

If F is universal, $F_1 \equiv k \pmod{2^{2e}}$ is solvable when k is arbitrary. Conversely, let there be solutions. If y is divisible by 2^e, the terms in y drop out and there is a solution with $y = 2^e$. In every case we may write $y = 2^s\eta$, η odd, $s \leqq e$. The solution gives one of $F_1 \equiv k \pmod{2^{e+s}}$. Insertion of this y into (3) gives a form F' which is derived from F by replacing y by η, e by $e+s$, b by $2^{2s}b$, and c by $2^s c$. Since F' has the same g, a, d, h, j, l as F, F' satisfies the initial conditions in §15. While N is unaltered, R and J are multiplied by 2^{2s}. Hence conditions (25) are unaltered. This proves that, *except in cases* (25), *F is universal if and only if $F_1 \equiv k \pmod{2^{2e}}$ is solvable when k is arbitrary*. Solutions with y even are here not excluded.

(I) $\lambda \equiv 3 \pmod 4$. The case $M \equiv 0 \pmod 4$ is excluded since $Z^2 + 3W^2$ takes only the values 0, 1, 3 modulo 4.

Let $M \not\equiv 0 \pmod 4$. Then F_1 with $x = 0$ represents all residues of 8. For $Z^2 + \lambda W^2$ represents exclusively 0, 1, 3, 4, 5, 7 (mod 8). The missing 2 and 6 are obtained from $y = 1$. We may select u from the six so that $M + u \equiv 2 \pmod 8$ since $2 - M \equiv 2$ or 6 only if $M \equiv 0$ or $-4 \pmod 8$. Similarly, we may select v from the six so that $M + v \equiv 6 \pmod 8$.

To proceed by induction from $m \geqq 3$ to $m+1$, let

$$x = \theta, \quad My^2 + Z^2 + \lambda W^2 = k + 2^m Q. \tag{29}$$

Then

$$M(y + 2^{m-1}\eta)^2 + (Z + 2^{m-1}\zeta)^2 + \lambda(W + 2^{m-1}\omega)^2 \equiv k \pmod{2^{m+1}}$$

if $Q + My\eta + Z\zeta + \lambda W\omega \equiv 0 \pmod 2$. The latter has solutions η, ζ, ω unless My, Z, W are all even. This disposes of odd k's.

Let $k \equiv 2 \pmod 4$. First, let M be odd and take $x = 0$. Then $My^2 + Z^2 + 3W^2 \equiv 2 \pmod 4$ shows that one of y, Z, W is even and two are odd. For y and Z odd, W even, $F_1 \equiv M+1$ or $M+5 \pmod 8$. For y and W odd, Z even, the values of F_1 are $M+\lambda$ and $M+\lambda+4$, i.e., $M+3$ and $M+7 \pmod 8$. Hence F_1 takes all even residues and therefore the value k modulo 8, when y is odd. By the above induction, $F_1 \equiv k \pmod{2^n}$ is solvable.

Second, let $k \equiv M \equiv 2 \pmod 4$. Then $y \equiv 1$, $Z \equiv W \pmod 2$. Take $x = 0$. For y, Z, W all odd, $F_1 \equiv M + 1 + \lambda \pmod 8$. Now $M + 1 + \lambda$ and k are congruent modulo 4. If they are congruent modulo 8, the preceding induction yields solutions modulo 2^n. There remains the case $k \equiv M + 1 + \lambda + 4 \pmod 8$. Write $k = 2\kappa$, $M = 2\mu$, $\lambda = 4t + 3$. Then κ and μ are odd and $\kappa \equiv \mu + 2t \pmod 4$. Since Z and W must now be even, write $Z = 2\zeta$, $W = 2\omega$. Thus $F_1 \equiv k \pmod{2^n}$ becomes

$$\mu y^2 + 2\zeta^2 + 2\lambda\omega^2 \equiv \kappa \pmod{2^{n-1}}, \quad y \text{ odd}. \tag{30}$$

Since $\kappa=\mu+2t+4s$, this holds modulo 8 if and only if

$$\zeta^2+3\omega^2\equiv t+2s \pmod 4.$$

This is solvable except when

$$(31)\qquad \begin{aligned} t+2s&\equiv 2 \pmod 4, & \kappa&\equiv \mu+4 \pmod 8,\\ k&\equiv M+8 \pmod{16}, & \lambda&\equiv 3 \pmod 8.\end{aligned}$$

In the latter case, $F_1\equiv k \pmod{16}$ has no solution with $x=0$ and hence no solution if $e\geqq 4$. This excludes the case $\lambda\equiv 3 \pmod 8$, $e\geqq 4$.

When (31_1) fails, (30) is solvable modulo 8. We proceed by induction. If (30) holds modulo 2^m, $m\geqq 3$, it holds modulo 2^{m+1} for the same ζ, ω, but with y replaced by $y+2^{m-1}\eta$, where η is determined modulo 2 since μy is odd. Hence $F_1\equiv k \pmod{2^n}$ is solvable when $k\equiv M\equiv 2 \pmod 4$, $\lambda\equiv 7 \pmod 8$.

Let (31) hold and $e=3$. We do not now take $x=0$. The conclusions preceding (30) continue to hold modulo 8. But (30) is now replaced by

$$(30')\qquad 4Axy+\mu y^2+2\zeta^2+2\lambda\omega^2\equiv\kappa \pmod{2^{n-1}},\quad y \text{ odd}.$$

This holds modulo 8 if $x\equiv Ay$, $\zeta\equiv w \pmod 2$. There is always a solution with $x=Ay$. Equate the left member to $\kappa+2^mQ$. Then if $m\geqq 3$,

$$\begin{aligned}(4A^2+\mu)(y+2^{m-1}\eta)^2+2\zeta^2+2\lambda\omega^2&\\ \equiv\kappa+2^m[Q+(4A^2+\mu)y\eta]\equiv\kappa &\pmod{2^{m+1}}\end{aligned}$$

by choice of η modulo 2.

Let (31) hold and $e=2$. The first coefficient 4 in $(30')$ is now replaced by 2. There is a solution with $x=2Ay$.

If in a solution of $F_1\equiv k \pmod{2^n}$ we multiply the variables by 2^s, we obtain a solution of $F_1\equiv 4^sk$.

(II) $\lambda\equiv 1 \pmod 4$. The case $M\equiv 0 \pmod 4$ is excluded since Z^2+W^2 takes only the values 0, 1, 2 modulo 4.

Modulo 8, $Z^2+\lambda W^2$ represents exclusively

$$(32)\qquad 0, 1, 4, 5, \lambda+1.$$

First, let M be odd. Then $My^2\equiv 0$, M, or $4M\equiv 4 \pmod 8$. Adding 4 to (31), we get the single new residue $\lambda+5$. It with $\lambda+1$ gives 2, 6 (mod 8). These with (32) give all residues except 3, 7. If one of the latter is congruent to the sum of M and a number (32), the latter must be even and hence 0, 4, or $\lambda+1$. These plus $M\equiv 1$ or 5 (mod 8) give a single new residue, viz., $\lambda+1+M$. If $e\geqq 3$, F_1 has a missing residue. Hence if $M\equiv 1 \pmod 4$ and $e\geqq 3$, F_1 has a missing residue. Hence if $M\equiv 1 \pmod 4$ and $e\geqq 3$, F_1 is not universal.

But if $e=2$, F_1 has the missing residue $\lambda+5+M$ modulo 8, when x, y, Z, W are all odd; and (28) represents all residues modulo 16.

If $M\equiv 3$ or 7 (mod 8), $M+0$ and $M+4$ give the missing 3, 7.

If $M\equiv 2$ (mod 4), $My^2\equiv 0$ or M (mod 8). The first four in (32) include all $\equiv 0$ or 1 (mod 4). Adding 0 and M, we get all residues modulo 4 and hence all modulo 8.

Hence if $M\equiv 2$ or 3 (mod 4), $F_1\equiv k$ is solvable modulo 8 with $x=0$ for every k. The first induction under (I) yields solutions modulo 2^n for every odd k.

Let $M\equiv 3$, $k\equiv 2$ (mod 4). For* $y=2Y$, Z and W odd, $F_1\equiv 4Y^2+1+\lambda$ (mod 8). According as $1+\lambda\equiv k$ or $k+4$ (mod 8), $F_1\equiv k$ holds for $Y=0$ or 1. To proceed by induction, note that (29) implies

$$My^2+(Z+2^{m-1}\zeta)^2+\lambda W^2\equiv k+2^m(Q+Z\zeta)\equiv k \pmod{2^{m+1}},$$

by choice of ζ modulo 2. Hence if $M\equiv 3$ (mod 4), $F_1\equiv k \pmod{2^n}$ is solvable when k, n are arbitrary.

Let $M\equiv k\equiv 2$ (mod 4). Then $2y^2+Z^2+W^2\equiv 2$ (mod 4). There are only two possibilities. Either Z and W are odd, and $y=2Y$, whence $F_1\equiv k$ if $k\equiv 1+\lambda$ (mod 8), with induction to 2^n. Or $Z=2\zeta$, $W=2\omega$, and y is odd; then, for $x=0$, $F_1\equiv M+4\zeta^2+4\omega^2\equiv M$ or $M+4$ (mod 8), whence $F_1\equiv k$ (mod 8) is solvable. For $x=0$, we have (30), which for modulus 8 is equivalent to $\zeta^2+\omega^2\equiv d$ (mod 4), where $2d=\kappa-\mu$. This is solvable unless

$$(33)\qquad d\equiv 3 \pmod 4,\quad 6\equiv\kappa-\mu \pmod 8,\quad 12\equiv k-M \pmod{16}.$$

There was a solution with $y=2Y$ unless $k\equiv 1+\lambda+4$ (mod 8). If both fail,

$$(34)\qquad M\equiv 1+\lambda \pmod 8.$$

Except in this case, $F_1\equiv k \pmod{2^n}$ is solvable. When $e\geqq 4$ and (33) and (34) hold, we saw that $F_1\equiv k$ (mod 16) has no solution. If (33) and (34) hold and $e=3$, we have (30′), which holds modulo 8 if and only if $\zeta+\omega$ is odd, $x\equiv Ay$ (mod 2). The former induction on (30′) applies also here. Likewise when $e=2$.

THEOREM 10. *Let $e\geqq 2$, $c=2C$, $j=2s$, d, l, and $\lambda=g^2d^2(lh-s^2)$ be odd. Write $M=\lambda g^2dbl-g^2d^2l^2C^2$. Then F is universal, if and only if we exclude cases* (25) *and the following*:

$$M\equiv 0 \pmod 4;\quad \lambda\equiv 3 \pmod 8,\quad e\geqq 4;\qquad \lambda\equiv M\equiv 1 \pmod 4,\ e\geqq 3;$$
$$\lambda\equiv 1 \pmod 4,\quad M\equiv 1+\lambda \pmod 8,\quad e\geqq 4.$$

* These are the only possible values modulo 4.

18. Next, let L be odd and λ even in the notations of §17. It suffices to treat the form (27) which we denote by

$$f = 2^e Axy + By^2 + 2Ryz + \lambda z^2 + W^2 \qquad (e \geqq 2). \tag{35}$$

If $e=2$, $2R+\lambda \equiv 2 \pmod 4$, f is universal except in cases (25). For, when y is odd, $f-B$ has the values 0, 1, $\tau = 2R+\lambda$, $\tau+1 \pmod 4$, which form a complete set of residues modulo 4. If $e \geqq 3$, we separate the case $\tau \equiv 2$ into two subcases.

I. Let $e \geqq 2$, $\lambda \equiv 0 \pmod 4$, R odd. Assign a fixed odd value to y. Then $f - By^2 \equiv 2tz + 4hz^2 + W^2 \pmod{2^e}$, where $t = Ry$ is odd and $4h = \lambda$. By (I) of §9, $tz + 2hz^2$ ranges with z over a complete set of residues modulo 2^n. Take $n = e-1$. Hence $2tz + 4hz^2$ takes all even values modulo 2^e. By use of $W = 0$ and 1, we see that f takes all values modulo 2^e.

II. Next, let $e \geqq 3$, $\lambda = 2r$, $R = 2\rho$, where r is odd. Then

$$\begin{gathered} F_1 = rf = 2^e rAxy + Ty^2 + 2Z^2 + rW^2, \quad Z = rz + \rho y, \\ T = rB - 2\rho^2. \end{gathered} \tag{36}$$

Since y enters linearly only in the first term, the first italicized result in §17 holds here. Hence we ask if $F_1 \equiv k \pmod{2^{2e}}$ is solvable when k is arbitrary and both odd and even y's are allowed.

The values modulo 8 of $2Z^2 + rW^2$ (r odd) are

$$0, 2, 4, 6, r, r+2. \tag{37}$$

If $T \equiv 0 \pmod 8$, F_1 has only these six values and is excluded. If $T \equiv 2$ or $6 \pmod 8$, the values of F_1 are (37) and the same increased by 2 or 6, and hence are (37) and either $r+4$ or $r+6$; thus F_1 has only seven values and is excluded.

Let $T = 4t$, where t is odd, and consider even values of F_1. Then $W = 2w$, and F_1 is the double of

$$2^{e-1} rAxy + \phi, \qquad \phi = 2ty^2 + Z^2 + 2rw^2.$$

Since $2r \equiv \pm 2 \pmod 8$, $Z^2 + 2rw^2 \equiv 0, 1, 2, 4, 6, 1 \pm 2 \pmod 8$. The further values of ϕ are obtained from these by adding $2t$. If such sums yield both missing values 1 ∓ 2, 5, they are obtained by adding the odd 1, 1 ± 2 to $2t$. This is impossible since t is odd and $2t \equiv 2$ or $6 \pmod 8$. This excludes $e \geqq 4$. But F_1 yields universal forms if $e = 3$. First, if y is odd, the values of F_1 modulo 8 are (37) increased by 4 and hence are all even residues and $r+2$, $r+6$. We lack r, $r+8$, $r+2$, $r+10 \pmod{16}$. We get these odd residues by taking $y = 2Y$, Y odd, whence $F_1 \equiv 2Z^2 + rW^2$, whose odd values modulo 16 are derived by adding 0, 2, 8 to each r, $9r$. The sum $9r+2 \equiv r+10 \pmod{16}$.

Finally, let T be odd. Then F_1 represents all residues modulo 8, as shown

by the first four numbers (37) and the values 0 and 1 of y. If k is odd, $F_1 \equiv k \pmod{2^n}$ is solvable with $x=0$. For, if $F_1 = k+2^m Q$, then

$$T(y+2^{m-1}\eta)^2 + 2Z^2 + r(W+2^{m-1}w)^2 \equiv k + 2^m(Q + Ty\eta + rWw) \equiv k \pmod{2^{m+1}}$$

by choice of η, w modulo 2.

Henceforth, let $k=2\kappa$. Then $y+W$ is even. We may set $y=\eta+\zeta$, $W=\eta-\zeta$, $T+r=2M$. From $F_1 \equiv 2\kappa \pmod{2^n}$, we cancel 2 and get

$$2^{e-1}rAxy + M\eta^2 + M\zeta^2 + 2(M-r)\eta\zeta + Z^2 \equiv \kappa \pmod{2^{n-1}}. \tag{38}$$

(a) If M is odd, multiply (38) by M and write

$$Y = M\eta + (M-r)\zeta, \qquad \lambda = 2Mr - r^2 \equiv 1 \pmod 4.$$

We get

$$2^{e-1}MrAxy + MZ^2 + Y^2 + \lambda\zeta^2 \equiv \kappa M \pmod{2^{n-1}}. \tag{39}$$

Aside from the first term, in which $y=\eta+\zeta$, this is of type (27) for the case (II) of §17 and M odd. It was proved there that, when $M \equiv 3 \pmod 4$, (39) is solvable with $x=0$ for every κ; but, when $M \equiv 1 \pmod 4$, $MZ^2+Y^2+\lambda\zeta^2$ represents all residues of 8 except $\lambda+5+M$. Hence if $e \geqq 4$, F_1 is not universal. The same is true if $e=3$ since

$$4MrAxy + MZ^2 + Y^2 + \lambda\zeta^2 \equiv \lambda + 5 + M \pmod 8, \quad y = \eta + \zeta, \tag{40}$$

are not solvable. For, $Z^2+Y^2+\zeta^2 \equiv 3 \pmod 4$ requires that Z, Y, ζ be all odd. By $Y \equiv \eta \pmod 2$, η is odd and y even. The left member of (40) is $M+1+\lambda \pmod 8$.

(b) Let $M=2m$. Write t for the odd integer $M-r$. Let ϕ denote the sum of the terms other than the first and last in the left member of (38). Then $\phi = 2m\eta^2+2m\zeta^2+2t\eta\zeta$. For η even, $\phi \equiv 0, 2m, 2m+4 \pmod 8$. By symmetry, ϕ takes the same values when ζ is even. For η and ζ both odd, $\phi \equiv 4m \pm 2 \pmod 8$.

(b_1) Let $m=2\mu$. Then ϕ takes all even values modulo 8. This holds also modulo 2^e. For proof, take $\eta=1$. Then $\phi = 2m+2\psi$, $\psi = 2\mu\zeta^2+t\zeta$. By (I) of §9, ψ ranges with ζ over a complete set of residues modulo 2^{e-1}. Give Z the values 0 and 1. Hence (38) is solvable with $x=0$, κ arbitrary.

(b_2) Let m be odd. Then $\phi \equiv 0, 2, 6 \pmod 8$. Hence (38) is not always solvable modulo 8 if $e \geqq 4$ and F_1 is then not universal. This is true also if $e=3$. For, we saw that $\phi+Z^2$ fails to represent 4 or 5 (mod 8). Suppose that

$$4rAxy + \phi + Z^2 \equiv 4 \text{ or } 5 \pmod 8. \tag{41}$$

Then

$$\phi + Z^2 \equiv 0 \text{ or } 1, \quad \phi \not\equiv 2, \quad \phi \equiv 0 \pmod 4, \quad \eta^2 + \zeta^2 + \eta\zeta \equiv 0 \pmod 2,$$

whence η and ζ are even. Thus y is even and (41) is impossible.

THEOREM 11. *Let $e \geqq 2$, $c = 2C$, $j = 2s$, d and l be odd. Write $\lambda = g^2d^2 \cdot (lh - s^2)$, $R = gdlC$. If $\lambda \equiv 0 \pmod 4$, F is universal. Next, let $\lambda = 2r$, $R = 2\rho$, where r is odd. If $e = 2$, F is universal. For $e \geqq 3$, write $T = rg^2dlb - 2\rho^2$. If $T \equiv 0$, 2, or 6 (mod 8), F is not universal. If $T \equiv 4 \pmod 8$, F is universal if and only if $e = 3$. If T is odd, write $2M = T + r$. Then F is universal if and only if $M \equiv 0$ or 3 (mod 4). Throughout, universality is to be qualified by excepting cases* (25).

19. Consider (35) for R odd, $\lambda \equiv 2 \pmod 4$. We may assume that $R \equiv 1 \pmod{2^{e+2}}$. For, if $R = 1 + 2\gamma$, $\lambda = 2r$, replace z by $z + \epsilon y$. We obtain a form like (35) with

$$R' = 1 + 2(\gamma + r\epsilon), \qquad B' = B + 2R\epsilon + 2r\epsilon^2.$$

Since r is odd, we may choose ϵ so that $\gamma + r\epsilon \equiv 0 \pmod{2^{e+1}}$. Note that $B' \equiv B \pmod 4$, so that the later conditions for universality are independent of this transformation.

If $B \equiv 1 \pmod 4$, there is no solution of $f \equiv \lambda + 5 \pmod 8$. The latter implies

$$(y + z)^2 + z^2 + W^2 \equiv 3 \pmod 4.$$

Whence $y + z$, z, W are all odd. Then $y = 2Y$ and

$$f \equiv 4Y^2 + 4Y + \lambda + 1 \equiv \lambda + 1 \pmod 8.$$

If $B \equiv 2 \pmod 4$, there is no solution of $f \equiv 5 \pmod 8$. That implies $2\phi + W^2 \equiv 1 \pmod 4$, $\phi = y^2 + yz + z^2$, whence W is odd and ϕ is even. Thus y and z are even and $f \equiv W^2 \pmod 8$.

Hence when f is universal, $B \equiv 0$ or 3 (mod 4).

First, let $x = 0$, $y = 1$, $z = 2Z$. Then $f = B + 4Z + 4\lambda Z^2 + W^2$. By (I) of §9, $Z + \lambda Z^2$ ranges with Z over a complete set of residues modulo 2^n. Hence f represents all $4t + B$ and $4t + B + 1 \pmod{2^n}$.

Second, let $x = 0$, $y = 2Y$, $z = 1$, where Y is odd. Then

$$f \equiv 4BY^2 + 4Y + \lambda + W^2. \tag{42}$$

If k is any odd integer, $f \equiv 2k \pmod{2^n}$ is solvable with $W = 2\omega$. Since $2k - \lambda$ is a multiple 4ϵ of 4, this is equivalent to $BY^2 + Y + \omega^2 \equiv \epsilon \pmod{2^{n-2}}$. But $Y = 2\eta + 1$. Hence $2\tau + B + 1 + \omega^2 \equiv \epsilon$, where $\tau = 2B\eta^2 + (2B + 1)\eta$. But τ ranges with η over a complete set of residues. Hence the congruence

is solvable with $\omega=0$ or 1. Thus if κ is any integer $\equiv 2 \pmod 4$, $f\equiv\kappa \pmod{2^n}$ is solvable with y a double of an odd integer.

Third, let $x=0$, $z=1$, $y=4\eta$, $\eta=2\xi+1$. Then

$$f \equiv 16B\eta^2 + 8\eta + \lambda + W^2 = 16\mu + 16B + 8 + \lambda + W^2,$$

where $\mu=4B\xi^2+(4B+1)\xi$ ranges with ξ over a complete set of residues modulo 2^n. Take $W=1$ and 3. Thus if $k\equiv\lambda+1 \pmod 8$, $f\equiv k \pmod{2^n}$ is solvable.

When $B\equiv 0 \pmod 4$, our three cases dispose of all except $\lambda+5 \pmod 8$. We then take $W=1$ in (42). Cancellation of 4 now gives $BY^2+Y\equiv 1$, $2\tau+B+1\equiv 1 \pmod{2^{n-2}}$, which is solvable for η.

When $B\equiv 3 \pmod 4$, our first two cases dispose of all except 1 (mod 4). Take $x=0$, $y=4\eta=4(2\xi+1)$, $z=2$. Then

$$f = 16B\eta^2 + 16\eta + 4\lambda + W^2 = 32u + 16(B+1) + 4\lambda + W^2,$$

where $u=2B\xi^2+(2B+1)\xi$ ranges with ξ over a complete set of residues modulo 2^n. Take $W=1, 3, 5, 7$. Then $32u+W^2$ represents all $8\sigma+1$, and the same is true of f modulo 2^n. There remains only the $8\sigma+5$. We take $x=0$, $z=2$, $y=2Y$, $Y=2\eta+1$. Write $B+1=4b$. Then

$$Bf = 4\epsilon^2 - 4 + 4B\lambda + BW^2, \quad \epsilon = BY + 1 = 2\tau,$$
$$\tau = B\eta + 2b.$$

We may employ τ as a new variable in place of η modulo 2^n. If $k\equiv B \pmod 8$, $16\tau^2+BW^2\equiv k \pmod{2^n}$ is solvable with W odd. For $m\geqq 3$, let $16\tau^2+BW^2 = k+2^mQ$. Then

$$16\tau^2 + B(W + 2^{m-1}\omega)^2 \equiv k + 2^m(Q + BW\omega) \equiv k \pmod{2^{m+1}}$$

by choice of ω modulo 2. Hence if $\kappa\equiv B+4 \pmod 8$, $Bf\equiv\kappa \pmod{2^n}$ is solvable. Thus $f\equiv 8\sigma+5 \pmod{2^n}$ is solvable when σ is arbitrary.

THEOREM 12. *Employ the initial assumptions and notations of Theorem* 11. *Write* $B=g^2dbl$. *Now let* R *be odd and* $\lambda\equiv 2 \pmod 4$. *Except in cases* (25), F *is universal if and only if* $B\equiv 0$ *or* 3 (mod 4).

20. Finally, consider (35) for $R=2\rho$, $\lambda=4r$. Then $f\equiv By^2+W^2 \pmod 4$, and f has the values 0, 1, B, $B+1$, which form a complete set of residues only if $B\equiv 2 \pmod 4$.

First, let $e=2$. For y odd, $f\equiv 2$ or 3 (mod 4). For $y=2Y$, $f\equiv 4rz^2+W^2 \pmod 8$, independently of Y. If r is even, $f\equiv 5 \pmod 8$ is impossible, and F is not universal. But if r is odd, $f\equiv 0, 1, 4, 5 \pmod 8$ for Y odd, and F is universal.

Next, let $e \geqq 3$. If r is even, $f \equiv By^2+4\rho yz+W^2 \equiv 5 \pmod 8$ is impossible, since it holds modulo 4 only when W is odd and y is even. If r and ρ are odd, $f \equiv By^2+4yz+4z^2+W^2 \equiv B+5 \pmod 8$ is impossible, since it holds modulo 4 only when W and y are odd.

There remains only the case $B=2\beta$, $\rho=2\sigma$, with β, σ, r all odd. Write $Z=rz+\sigma y$, $T=r\beta-2\sigma^2$. Then

$$F_1 = rf = 2^e rA\,xy + \phi, \quad \phi = 4Z^2 + 2Ty^2 + rW^2.$$

As in §17, F is universal except in cases (25) if and only if $F_1 \equiv k \pmod{2^{2e}}$ is solvable when k is arbitrary, when both even and odd values of y are allowed. The even residues modulo 8 of ϕ are 0, 4, $2T$, $2T+4$, which are a permutation of 0, 2, 4, 6. The odd residues are obtained by adding r to these four. Hence $F_1 \equiv k \pmod 8$ is solvable when k is arbitrary.

For κ odd, $\kappa - T = 2q$. Then $\phi \equiv 2k \pmod{16}$ requires

$$W = 2w, \quad y \text{ odd}, \quad Z^2 + rw^2 \equiv q \pmod 4.$$

But $Z^2+rw^2 \equiv 0, 1, r, r+1 \pmod 4$, which lack 3 or 2 according as $r \equiv 1$ or $3 \pmod 4$. Hence $\phi \equiv k \pmod{16}$ is impossible for certain integers k, whence F is not universal if $e \geqq 4$.

If $e=3$, $B \equiv 2 \pmod 4$, r is odd and ρ even, we shall prove that F is universal. For y odd,

$$f - B \equiv 4z^2 + W^2 \equiv 0,1,4,5 \pmod 8.$$

Since $f-B$ takes the values 0, 1 (mod 4),

$$f \equiv 2,3,6,7,10,11,14,15 \pmod{16}. \tag{43}$$

For $y=2Y$, Y odd, $f-4B \equiv \psi = 4rz^2+W^2 \pmod{16}$.

(i) Let $r \equiv 1 \pmod 4$. Then $\psi \equiv 0, 1, 4, 5, 8, 9, 13 \pmod{16}$. To these we add $4B \equiv 8$ and get 0, 1, 5, 8, 9, 12, 13 (mod 16). From these and (43) only 4 is missing. For $y=4\eta$, $f \equiv 4rz^2+W^2 \pmod{32}$. We take $W=2$, $z=0$, 2 and get 4, 20 (mod 32) and hence the missing 4 (mod 16).

(ii) Let $r \equiv 3 \pmod 4$. Then $\psi \equiv 0, 1, 4, 5, 9, 12, 13 \pmod{16}$. To these we add $4B \equiv 8$ and get 1, 4, 5, 8, 9, 12, 13. From these and (43) only 0 is missing. We use $y=4\eta$, $z=0$, $W=0, 4$ and get 0, 16 (mod 32).

THEOREM 13. *Employ the initial assumptions and notations of Theorem* 11. *Write* $B=g^2dbl$. *Now let* $R=2\rho$, $\lambda=4r$. *Except in cases* (25), F *is universal if and only if* $B \equiv 2 \pmod 4$, *either* $e=2$, *and* r *odd, or* $e=3$, r *odd and* ρ *even.*

Theorems 10–13 cover all classic forms with $e>1$, d and l odd, c and j even. The remaining forms will be treated in a later paper.

PART III. EQUIVALENCE AND CANONICAL FORMS

21. We consider henceforth only classic forms f in which the coefficients of products of different variables are all even. By the Hessian of f is meant the determinant of the halves of the second partial derivatives of f. We shall prove

THEOREM 14. *The Hessian H of a universal classic ternary quadratic Null form f is either odd or the double of an odd integer. In the respective cases, f is equivalent to*

$$\phi = 2xy - Hz^2, \quad \psi = 2xy + y^2 - Hz^2. \tag{44}$$

Each of these forms is universal.

By §8, f is equivalent to

$$2Axy + By^2 + 2Cyz + Dz^2, \quad A \text{ odd}. \tag{45}$$

First, let $A=1$. Replacing x by $x+ky-Cz$, we get a form like (45) with $C'=0$, $B'=B+2k=0$ or 1 by choice of k. Then $H=-D$ and we have (44).

First, let H be even. Then ϕ is excluded. If $H\equiv 0 \pmod 4$, ψ is never $\equiv 2 \pmod 4$. Hence $H\equiv 2 \pmod 4$ in ψ. For $y=1, z=0$; $y=2, z=0, 1$, the values of ψ are $2x+1$, $4(x+1)$, $4(x+1)-H$, which together give all integers.

Second, let H be odd. In ψ replace x by $x+Hz$ and z by $z+y$; we get $2xy+(1-H)y^2-Hz^2$. We now replace x by $x+\frac{1}{2}(H-1)y$ and get ϕ. Hence we may drop ψ. The values of ϕ for $y=1, z=0, 1$ are $2x$ and $2x-H$ and together give all integers.

This proves Theorem 14 when $A=1$.

22. **Case A prime to D.** There exist solutions s and t of $Ds-At=C$. Replacing x by $x+tz$, we get a form (45) having $C'=C+At=Ds$. We now, replace z by $z-sy$ and obtain a form (45) lacking yz. In the notation (3), it has $g=1$ and is

$$F = 2axy + by^2 + dz^2 \quad (a \text{ odd and prime to } d). \tag{46}$$

If F is universal the first part of §5 with $t=a$, $c=0$, shows that $-bd$ is a quadratic residue of each power of a prime which divides a and hence of a itself. Thus b is prime to a, and $b=-r^2d+am$, where r and m are integers, and r is prime to a.

(I) Let d be odd. If m is even, write $m=2\mu$, $r=\rho$. Then

$$b = -\rho^2 d + 2a\mu \quad (\rho, a \text{ relatively prime}). \tag{47}$$

If m is odd, then $m+2rd+ad$ is an even integer 2μ; for $\rho=r+a$ we again have (47). For $x=X-\mu y$, (46) becomes

$$F = 2aXy - \rho^2 dy^2 + dz^2.$$

Put $z=Z+\rho y$. Then $F=2y\xi+dZ^2$, where $\xi=aX+\rho dZ$. Since a and $d\rho$ are relatively prime, there exist solutions of $av-d\rho u=1$. Write $\eta=uX+vZ$. The transformation from X, Z to ξ, η is of determinant unity and replaces F by $2y\xi+d(a\eta-u\xi)^2$. Replacing y, ξ, η by x, y, z, we obtain a form (45) having $A=1$.

(II) Let d be even. If F is universal, b is odd and $d\equiv 2 \pmod 4$. For, if $d\equiv 0 \pmod 4$, F is never $\equiv 2 \pmod 4$. Thus $d=2D$, where D is odd. Then m is odd, $m=2n+1$. Replacing x by $x-ny$, we obtain

$$F = 2axy + Ty^2 + 2Dz^2, \quad T = a - 2r^2D.$$

Write $\theta=4r^2D-a$. Since $4rD$ is prime to a, it is prime to θ. Hence there exist integers β and γ satisfying $4rD\gamma-\theta\beta=1$. Now $F=0$ for $x=-1$, $y=2$, $z=2r$. Hence the substitution

$$x = -\xi, \quad y = 2\xi + \beta\eta + 4D\zeta, \quad z = 2r\xi + \gamma\eta + \theta\zeta$$

has determinant unity and replaces F by a form G in which the coefficient of ξ^2 is zero. That of $2\xi\eta$ is

$$-a\beta + 2T\beta + 4Dr\gamma = a\beta + 4rD(\gamma - r\beta) = 4rD\gamma - \theta\beta = 1.$$

That of $\xi\zeta$ is $8rD(\theta-a+2T)=0$. Replacing ξ, η, ζ by x, y, z, we see that G is of the type (45) with $A=1$.

The Hessian $-a^2d$ of (46) is odd in (I) and $\equiv 2 \pmod 4$ in (II). This completes the proof of

THEOREM 15. *When A is odd and prime to D,* (45) *is equivalent to* (46), *which is universal if and only if $-bd$ is a quadratic residue of a and either d is odd or b is odd and $d\equiv 2 \pmod 4$. In the respective cases,* (46) *is equivalent to* (44).

This proves Theorem 14 when A is prime to D.

23. Let the form (3) be universal and classic, whence $e\geqq 1$, $c=2C$, $\psi=c_{11}z^2+2c_{12}zw+\cdots$, where c_{11}, c_{12}, $\cdots$ have no common factor >1.

If b is even, then d is odd and c_{11}, c_{22}, c_{33}, $\cdots$ are not all even. We may assume that c_{11} is odd. If the number n of variables is 3, $c_{11}=1$. If $n>3$ and if c_{11} is even and c_{22}, for example, is odd, replace w by $w+z$; then the new c_{11} is odd. In (3) we replace z by $z+gy$ and obtain a form $2^e gaxy+\phi(y,z,w,\cdots)$ in which the coefficient of y^2 is the product of g by $b+2C+dg^2c_{11}$, which is odd. Hence (3) is equivalent to a like form with b odd.

Let D be the g.c.d. of $a=D\alpha$ and $b=D\beta$. Then D, α, β are odd, while

$2^e\alpha$ and β are relatively prime. Let E be the g.c.d. of $D=E\delta$ and $C=E\gamma$. Then δg is prime to 2γ. Hence there are solutions $r, s; M, v; h, u$ of

$$2^e\alpha r + \beta S = -2\gamma, \tag{48}$$

$$\delta g M + 2\gamma v = 1, \quad 2^e\alpha h + \beta u = M. \tag{49}$$

The form (3) is

$$F = 2^e gE\delta\alpha xy + gE\delta\beta y^2 + 2E\gamma yz + gd\psi(z, w, \cdots). \tag{50}$$

Consider the linear substitution

$$S: \quad \begin{aligned} x &= \beta X + hY + rZ, \\ y &= -2^e\alpha X + uY + sZ, \\ z &= vY + g\delta Z, \end{aligned}$$

which does not alter w, etc. Multiply the second row of its determinant Δ by β. To the new second row add the product of the first row by $2^e\alpha$ and apply (48) and (49_2). Hence

$$\beta\Delta = \beta \begin{vmatrix} M & -2\gamma \\ v & g\delta \end{vmatrix}, \qquad \Delta = 1$$

by (49_1). We see that S replaces F by a form in which the coefficient of X^2 is zero, and that of XY is

$$2^e gE\delta\alpha(\beta u - 2^e\alpha h) + gE\delta\beta(-2\cdot 2^e\alpha u) + 2E\gamma(-2^e\alpha v)$$
$$= -2^e\alpha E(g\delta M + 2\gamma v) = -2^e\alpha E,$$

by (49). The coefficient of XZ is the product of $-2^e gE\alpha\delta$ by $-\beta S + 2^e\alpha r + 2\beta s + 2\gamma = 0$, by (48). Changing the signs of Y and Z, we get an equivalent form

$$2^e\alpha EXY + \chi(Y, Z, w, \cdots). \tag{51}$$

Let m denote the minimum odd positive integer such that a given universal classic form is equivalent to $2^e mxy + \phi(y, z, w, \cdots)$. By (50), $m = gE\delta\alpha$. Since F is equivalent to (51), $\alpha E \geqq m$. Hence $1 \geqq g\delta$, $g = \delta = 1$.

Since $g=1$ and a is prime to d, when we express a ternary form (3) with $e=1$ in the notation (45), we see that A is prime to D. Hence Theorem 15 applies and proves Theorem 14.

Incidentally we have also

THEOREM 16. *Every universal classic Null quadratic form is equivalent to* (2) *with* $A = 2^e m$, *where the minimum* m *is prime to* Δ, $c = 2C$, m *and* B *are odd and their g. c. d. divides* C.

PART IV. ALL UNIVERSAL $f=ax^2+by^2+cz^2$

24. THEOREM 17. *f is universal if and only if* (i) *a, b, c are not all of like sign and no one is zero*; (ii) *a, b, c are relatively prime in pairs*; (iii) *$-bc$, $-ac$, $-ab$ are quadratic residues of a, b, c, respectively*; (iv) *abc is odd or double an odd integer.*

This was proved elsewhere* by the writer. We shall give here a new proof that the conditions are sufficient. By† (i), (ii), (iii), $f=0$ has solutions ξ, η, ζ, relatively prime in pairs. Then $s\xi-r\eta=1$ has solutions s, r. The substitution

$$x = \xi X + rY, \quad y = \eta X + sY, \quad z = \zeta X + Z$$

has determinant unity and replaces f by

$$F = 2uXY + 2c\zeta XZ + vY^2 + cZ^2,$$

where $u = a\xi r + b\eta s$, $v = ar^2 + bs^2$.

The Hessians abc and $-c(u^2+cv\zeta^2)$ of f and F must be equal, whence

$$u^2 + cv\zeta^2 = -ab. \tag{52}$$

This follows also from the identity

$$u^2 - v(a\xi^2 + b\eta^2) \equiv -ab(s\xi - r\eta)^2$$

in ξ, η. By (52) and (ii), no prime divides both u and c. Suppose u and ζ have a common prime factor p. By (52), p divides ab. If p divides a, it divides the second term of $a\xi^2+b\eta^2+c\zeta^2=0$, whereas b is prime to a, and η to ζ. Similarly, p cannot divide b.

Hence the coefficients of $y_1=uY+c\zeta Z$ are relatively prime. Thus there is another linear function z_1 of Y and Z such that the determinant of the coefficients in y_1 and z_1 is unity. Hence F is equivalent to $2Xy_1+\phi$, where ϕ is quadratic in y_1 and z_1. The new form is of type (45) with $A=1$. Its Hessian abc has property (iv). By §21, it is universal.

Finally, we give a new proof that (i)-(iv) are necessary conditions that a Null form f be universal. Such an f is equivalent to F in §8 with $e=1$. Then the Hessian of F is $-A^2D$, where D is odd or double an odd. This proves (iv). If a and b have a common odd prime factor p, $f\equiv cz^2 \pmod p$ and f would not be universal. This proves (ii). Suppose that $-bc$ is a non-residue of an odd prime factor p of $a=pA$. Consider values x, y, z for which f is divisible by p. Then $-bcy^2\equiv(cz)^2 \pmod p$, whence $y\equiv z\equiv 0$, $f=pF$, $F\equiv Ax^2 \pmod p$, and f would not be universal.

* Bulletin of the American Mathematical Society, vol. 35 (1929).

† Dirichlet-Dedekind, *Zahlentheorie*, 4th edition, §157, p. 432.

UNIVERSITY OF CHICAGO,
CHICAGO, ILL.

BOUNDARY VALUE PROBLEMS FOR POTENTIALS OF A SINGLE LAYER*

BY

E. R. C. MILES

1. Introduction. The principal object of the following paper is the discussion of a Neumann problem, with reference to a potential of a single layer which is based on a general distribution of matter on a simple closed plane boundary. Such potentials were introduced by Plemelj.† The result obtained here is of the same order of generality for these boundaries as that obtained by G. C. Evans with the aid of conformal transformations‡, but the present method is entirely different, and simpler. The problem is equivalent to a Stieltjes integral equation, which is solved by reduction to the classical Fredholm type.

Consider the function

$$v(M) = \int_C \log (1/r)d\mu(P), \tag{1}$$

where $\mu(P)$ is a function of limited variation on the simple closed curve C. The notation $\mu(P)$ is used as synonymous with $\mu(s_P)$. For the measurement of angles the length $r = MP$ is regarded as directed from M to P, where M is a point which remains fixed in the integration. It is assumed that the direction of C changes continuously as we go along the curve, and that the curvature exists at each point.

Let n, or, more particularly, n_Q, be the normal at a point Q of the curve, directed toward the interior, and M a point on n_Q. The derivative

$$\frac{dv(M)}{dn} = \int_C \frac{\cos (n_Q, r)}{r} d\mu(P) \tag{2}$$

generally fails to be continuous as M passes over the point Q. We shall investigate this discontinuity.

The integral (2) converges when M is coincident with Q. In order to prove this statement, it suffices to examine the integrand in the neighborhood

---

* Presented to the Society, April 7, 1928; received by the editors June 2, 1928.

† *Potentialtheoretische Untersuchungen*. Leipzig, 1911, p. 17.

‡ *The Logarithmic Potential. Discontinuous Dirichlet and Neumann Problems*, New York, 1927, p. 84.

of Q. Set up a system of rectangular coördinates, with origin at Q, and axes along the positive tangent and inner normal, respectively. Measure s from Q. In the neighborhood of Q, the equation of C may be written in the form

$$y = x^2 f(x),$$

where, as may be verified from the hypotheses for C, $f(x)$ and $xf'(x)$ are continuous except at $x=0$, and bounded in a suitable interval containing Q.

Denote by θ the angle measured in the positive direction from the normal at Q to QP. We have then

$$\frac{\cos\theta}{r} = \frac{y}{r^2} = \frac{f(x)}{1 + x^2[f(x)]^2},$$

which is therefore continuous except at Q as a function of x, and bounded in the neighborhood of Q. Hence it has the same properties as a function of s_P when $P \neq Q$. The statement is therefore proved.

From what has been said of the function $(\cos\theta)/r$, we obtain immediately the following corollary, which will be useful later. We state the result as a lemma.

LEMMA. *If $\mu(P)$ is continuous at Q, and C' is an arc of C, containing Q, and of length $<\delta$, then*

$$\int_C \frac{\cos\theta}{r} d\mu(P) = \lim_{\delta \to 0} \int_{C-C'} \frac{\cos\theta}{r} d\mu(P).$$

2. **Boundary values of the normal derivative.** For convenience we denote by $dv(Q+)/dn$ the value, where it exists, of dv/dn as M approaches Q along n from the inside; we define similarly $dv(Q-)/dn$ with reference to the exterior of C. With this notation, we have the following theorem:

THEOREM 1. *If $v(M)$ is given by* (1), *then*

$$\frac{dv(Q+)}{dn} + \frac{dv(Q-)}{dn} = 2\int_C \frac{\cos\,(n_Q, QP)}{QP} d\mu(P),$$

$$\frac{dv(Q+)}{dn} - \frac{dv(Q-)}{dn} = -2\pi\mu'(Q),$$

provided $\mu'(Q)$ exists.

Denote by C' a small arc HK containing Q as an interior point. We may write (2) in the form

$$\frac{dv(M)}{dn} = \int_{C-C'} + \int_{C'} \frac{\cos\theta}{MP} d\mu(P) = I_1 + I_2.$$

Consider I_2. Construct the osculating circle $\overline{C}$ at Q. Let ρ be the distance from its center to M, $\overline{P}$ that point of $\overline{C}$ for which $x=x_P$. Let $\bar{r}=M\overline{P}$, and $\bar{\theta}$ the angle from n_Q to $M\overline{P}$, measured in the positive direction. We shall show that

$$\left|\int_{C'}\left[\frac{\cos\theta}{r}-\frac{\cos\bar{\theta}}{\bar{r}}\right]d\mu(P)\right|<\epsilon,$$

uniformly with respect to M, provided that the length of C' is small enough. For this purpose it suffices to show that $(\cos\theta)/r-(\cos\bar{\theta})/\bar{r}$ is bounded uniformly with respect to M. In fact, $\mu(P)$ is of limited variation in the neighborhood of Q, and continuous at Q. The conclusion is apparent if we write

$$\frac{\cos\theta}{r}-\frac{\cos\bar{\theta}}{\bar{r}}=\frac{(\bar{y}-y)\cos(\theta+\bar{\theta})}{r\bar{r}}$$

and note (a), that $\bar{y}-y$ is of order higher than the second with respect to x, and (b), that $|x|<r, \bar{r}$. The details are as follows:

$$\bar{y}=R-(R^2-x^2)^{1/2},$$

where R is the radius of curvature of C at Q. But the preceding equation may also be written

$$\bar{y}=R\left(1-\left(1-\frac{x^2}{R^2}\right)^{1/2}\right)$$

$$=\frac{x^2}{2R}+x^4\Phi(x).$$

Then

$$y-\bar{y}=x^2\left[f(x)-\frac{1}{2R}\right]+x^4\Phi(x),$$

where the expression in brackets is bounded. Moreover $\cos(\theta+\bar{\theta})$ is obviously continuous as a function of x, and uniformly bounded with respect to M. The statement is therefore proved.

Consider therefore the integral

$$I_2=\int_{\overline{C}'}\frac{\cos\bar{\theta}}{\bar{r}}d\mu(P)=\frac{d}{dn}\int_{\overline{C}'}\log\frac{1}{\bar{r}}d\mu(P)$$

$$=\frac{1}{2}\frac{\partial}{\partial\rho}\int_{\overline{C}'}\log(R^2+\rho^2-2R\rho\cos\psi)d\mu(P)$$

$$=-\frac{1}{2\rho}\int_{\overline{C}'}\frac{-2\rho^2+2R\rho\cos\psi}{R^2+\rho^2-2R\rho\cos\psi}d\mu(P),$$

where ψ is the angle subtended by $Q\overline{P}$ at the center O of the circle, and where ρ is the segment OM, directed from O to M. Hence

$$-I_2+\int_{\overline{C}'}\frac{1}{2\rho}\,d\mu(P)=\frac{1}{2\rho}\int_{\overline{C}'}\frac{R^2-\rho^2}{R^2+\rho^2-2R\rho\cos\psi}\,d\mu(P).$$

If the length of $\overline{C}'$ is sufficiently small, we have, therefore,

$$\left|I_2+\frac{1}{2\rho}\int_{\overline{C}'}\frac{R^2-\rho^2}{R^2+\rho^2-2R\rho\cos\psi}\,d\mu(P)\right|<\epsilon.$$

But*

$$\lim_{\rho\to R}\frac{1}{2\pi R}\int_{\overline{C}'}\frac{R^2-\rho^2}{R^2+\rho^2-2R\rho\cos\psi}\,d\mu(P)=\mu'(Q).$$

Hence, taking account of the lemma of §1,

$$\frac{dv(Q+)}{dn}=-\pi\mu'(Q)+\int_C\frac{\cos(n_Q,QP)}{QP}\,d\mu(P),$$

and similarly

$$\frac{dv(Q-)}{dn}=\pi\mu'(Q)+\int_C\frac{\cos\ (n_Q,QP)}{QP}\,d\mu(P).$$

From these equations we have the result stated in the theorem.

The foregoing analysis presupposes that the curvature at Q is not zero. In the latter case, we may replace the arc of the osculating circle by the corresponding segment $\overline{C}'$ of the tangent at Q, and observe that the integrals along C' and along $\overline{C}'$ differ by a quantity which is uniformly small with the length of C'. In fact, $(\cos\theta)/r-(\cos\bar{\theta})/\bar{r}$ is uniformly bounded with respect to M, since y is of order higher than the second with respect to x. Therefore consider the integral†

$$\int_{-x_0}^{x_0}\frac{\cos\bar{\theta}}{\bar{r}}\,d\mu(x)$$

$$=\mu'(Q)\int_{-x_0}^{x_0}\frac{\cos\bar{\theta}}{\bar{r}}\,dx-\int_{-x_0}^{0}\frac{\bar{y}}{\bar{r}^2}\,d(x\eta(x))-\int_0^{x_0}\frac{\bar{y}}{\bar{r}^2}\,d(x\eta(x)),$$

where $\mu(x)=x\mu'(Q)+x\eta(x)$, the last term being a function of x of limited variation, such that if η_0 denotes the upper bound of $|\eta(x)|$ in the interval $-x_0\leqq x\leqq x_0$, we have $\lim_{x_0\to 0}\eta_0=0$.

* Evans, loc. cit., pp. 39, 40.

† This integral is essentially the Poisson Stieltjes integral for the straight line considered as an infinite circle.

Now since $\bar{y}/\bar{r}^2$ decreases with x, $x>0$, we have

$$\int_0^{x_0} \frac{\bar{y}}{\bar{r}^2}\, d(x\eta(x)) = \frac{\bar{y}}{\bar{r}^2}\, x_0\eta(x_0) - \int_0^{x_0} x\eta(x) d\left(\frac{\bar{y}}{\bar{r}^2}\right),$$

where

$$\left|\frac{\bar{y}}{\bar{r}^2}\, x_0\eta(x_0)\right| < \eta_0,$$

and

$$\begin{aligned}\left|\int_0^{x_0} x\eta(x) d\left(\frac{\bar{y}}{\bar{r}^2}\right)\right| &\leqq -\eta_0 \int_0^{x_0} x d\left(\frac{\bar{y}}{\bar{r}^2}\right) \\ &\leqq \eta_0 x_0 \left[\frac{\bar{y}}{\bar{r}^2}\right]_{x=x_0} + \eta_0 \int_0^{x_0} \frac{\bar{y}}{\bar{r}^2}\, dx \\ &\leqq \eta_0\left(1+\frac{\pi}{2}\right).\end{aligned}$$

Hence the integral considered is made arbitrarily small with x_0, uniformly for M. The same remark applies to

$$\int_{-x_0}^{0} \frac{\bar{y}}{\bar{r}^2} d(x\eta(x)).$$

But $\int_{-x_0}^{x_0} (\cos\bar{\theta})/\bar{r}\, dx = 2(\theta_0 - \pi)$. Hence, by first taking x small enough, and then taking M near enough to Q, we can make

$$\int_{-x_0}^{x_0} \frac{\cos\bar{\theta}}{\bar{r}} d\mu(x)$$

differ by as little as we please from $\pi\mu'(Q)$, which is what we wished to show.

3. **The Stieltjes integral equation.** We consider again the class of functions which can be expressed in the form (1). It is assumed now that the discontinuities of $\mu(P)$ are regular. If we also denote by C the total length of the curve, we may extend the definition of $\mu(s_P)$ beyond $s_P = C$ by means of the relation

$$\mu(s_P + C) = \mu(s_P) + m,$$

where

$$m = \int_C d\mu(s_P) = \mu(C).$$

The problem now to be considered is expressed in terms of the condition

$$(3) \qquad \frac{1+\lambda}{2\lambda} \lim_{\delta\to 0+} \int_{A_\delta}^{B_\delta} \frac{dv}{dn} ds_\delta - \frac{1-\lambda}{2\lambda} \lim_{\delta\to 0-} \int_{A_\delta}^{B_\delta} \frac{dv}{dn} ds_\delta = g(B) - g(A),$$

where $g(s)$ is a given function of limited variation, with regular discontinuities, and where A, B are points of C. We shall regard A as fixed and B as variable on C. In the first integral s_δ is an arbitrary simple regular curve lying entirely within C, and joining the points A_δ, B_δ. The latter points are at the distance δ from A, B, measured along the respective inner normals. In the second integral s_δ lies entirely outside C; the points A_δ, B_δ are on the outer normals at A, B, respectively.

That the curve C has a unique normal at each point is implied in the preceding statements. Specifically, it is assumed that C is a simple regular closed curve; moreover C shall be a curve of class Γ,* i.e., there shall exist a positive number Γ, independent of M, such that

$$\int_C \frac{|\cos(n_P, MP)|}{MP} ds_P < \Gamma,$$

uniformly for all M.

THEOREM 2. *There is a unique function of class* (1) *which satisfies the condition* (3) *unless* λ *belongs to a certain set of characteristic values. These values are independent of the given function* $g(B)$, *and depend merely on the form of the curve* C.

The value $\lambda = -1$ *is not a characteristic value. The corresponding problem is a generalized Neumann problem for the exterior region*:

$$\lim_{\delta \to 0-} \int_{A_\delta}^{B_\delta} \frac{dv}{dn} ds_\delta = g(B) - g(A).$$

The value $\lambda = +1$ *is a characteristic value. A necessary and sufficient condition that there exist a solution* $v(M)$ *of class* (1) *is that*

$$\int_C dg(B) = g(C) = 0.$$

The corresponding problem is a generalized Neumann problem for the interior region:

$$\lim_{\delta \to 0+} \int_{A_\delta}^{B_\delta} \frac{dv}{dn} ds_\delta = g(B) - g(A).$$

The solution is unique except for an arbitrary additive constant.

Consider the condition (3). We have

$$\lim_{\delta \to 0+} \int_{A_\delta}^{B_\delta} \frac{dv}{dn} ds_\delta = -\lim_{\delta \to 0+} \int_{A_\delta}^{B_\delta} \frac{dw}{ds_\delta} ds_\delta = \lim_{\delta \to 0+} [w(A_\delta) - w(B_\delta)],$$

* G. C. Evans, The Rice Institute Pamphlet, vol. 7 (1920), p. 261.

where w is a function conjugate to v:

$$w(M) = -\int_C \tan^{-1}\frac{y_P - y_M}{x_P - x_M}\,d\mu(P).$$

The function under the integral sign being multiple valued, we write

$$w(M) = -\int_C \theta(M,P)d\mu(P),$$

where $\theta(M, P)$ is the angle from MA to MP measured in the positive direction. Then, taking A as the initial point of the integration, θ varies continuously as P traces the curve.

Integrate by parts:

$$\begin{aligned} w(M) &= -\left[\mu(P)\theta(M,P)\right]_A^A \\ &\quad + \int_C \mu(P)\frac{d\theta(M,P)}{ds_P}\,ds_P \\ &= \phi(M) + u(M). \end{aligned} \tag{4}$$

We shall take $\mu(A)=0$, since this can be effected by the addition of a constant. Then

$$\phi(M) = \begin{cases} -2\pi m, & M \text{ inside } C, \\ 0, & M \text{ outside } C. \end{cases}$$

The function $u(M)$ is a *double layer* potential with density a function of limited variation. The integral converges when M is a point on C, since $\mu(P)$ is bounded, and C is of class Γ.

Consider $u(Q+)$ as M approaches a point Q of C along the normal:

$$u(M) = \int_C \mu(P)\frac{d\theta(M,P)}{ds_P}\,ds_P = \int_C \mu(P)\frac{\cos(n_P, MP)}{MP}\,ds_P.$$

In order to study the behavior of $u(M)$ in the neighborhood of Q, we need to extend the traditional analysis to cover the case where $\mu(P)$ has a regular discontinuity at Q.

Write

$$u(M) = \int_{C-C'} + \int_H^Q + \int_Q^K \mu(P)\frac{\cos(n_P, MP)}{MP}\,ds_P,$$

where C' denotes a small arc HK, of which Q is an interior point.

The first of the preceding integrals is continuous at Q, since Q is not on C'. Then

$$\lim_{M\to Q}\int_{C-C'} \mu(P)\frac{\cos(n_P, MP)}{MP}\,ds_P = \int_{C-C'} \mu(P)\frac{\cos(n_P, QP)}{QP}\,ds_P.$$

The second integral may be written

$$\int_H^Q [\mu(P)-\mu(Q-)]\frac{\cos(n_P, MP)}{MP}ds_P+\mu(Q-)\int_H^Q \frac{\cos(n_P, MP)}{MP}ds_P = I_1+I_2,$$

where $\lim_{P\to Q}\mu(P)=\mu(Q-)$, the limit existing since $\mu(P)$ is of limited variation. Hence H may be taken near enough to Q so that $|\mu(P)-\mu(Q-)|<\epsilon/\Gamma$, where $\epsilon>0$ is arbitrarily small. By hypothesis, C is of class Γ. Consequently

$$|I_1|<\epsilon$$

if s_{HQ} is sufficiently small, and uniformly with regard to M.

Write I_2 in the form

$$I_2 = \mu(Q-)\int_H^Q \frac{d}{ds}\tan^{-1}\frac{y_M-y_P}{x_M-x_P}ds_P$$
$$= \mu(Q-)\left[\tan^{-1}\frac{y_M-y_Q}{x_M-x_Q}-\tan^{-1}\frac{y_M-y_H}{x_M-x_H}\right].$$

The quantity within brackets is the angle from MH to MQ, measured in the positive direction. As M approaches Q along the normal, this angle approaches as its limit the angle from the chord QH to the (outer) normal at Q. Hence if H is taken near enough to Q, the limiting angle may be made as near $\pi/2$ as we wish. That is,

$$\lim_{M\to Q+} I_2(M) = \frac{\pi}{2}\mu(Q-)+\eta(H),$$

where $\lim_{H\to Q}\eta(H)=0$.

By a similar process we obtain

$$\int_Q^K \mu(P)\frac{\cos(n_P, MP)}{MP}ds_P$$
$$= \int_Q^K [\mu(P)-\mu(Q+)]\frac{\cos(n_P, MP)}{MP}ds_P+\mu(Q+)\int_Q^K \frac{\cos(n_P, MP)}{MP}ds_P$$
$$= I_3+I_4.$$

By taking K sufficiently near to Q, $|I_3|$ can be made arbitrarily small, uniformly with respect to M. Also

$$\lim_{M\to Q+} I_4(M) = \frac{\pi}{2}\mu(Q+)+\eta'(K), \quad \lim_{K\to Q}\eta'(K)=0.$$

If M approaches Q along the outer normal at Q, the corresponding limits are given by

$$\lim_{M\to Q-} I_2'(M) = -\frac{\pi}{2}\mu(Q-) + \eta''(H),$$

$$\lim_{M\to Q-} I_4'(M) = -\frac{\pi}{2}\mu(Q+) + \eta'''(K),$$

$$\lim_{H\to Q} \eta''(H) = \lim_{K\to Q} \eta'''(K) = 0.$$

We have, therefore,

$$u(Q+) = \int_C \mu(P)\frac{\cos(n_P, QP)}{QP}ds_P + \frac{\pi}{2}[\mu(Q+) + \mu(Q-)]$$

$$= \int_C \mu(P)\frac{d}{ds_P}\tan^{-1}\frac{y_Q - y_P}{x_Q - x_P}ds_P + \pi\mu(Q),$$

since the discontinuities of $\mu(P)$ are regular. Similarly

$$u(Q-) = \int_C \mu(P)\frac{d}{ds_P}\tan^{-1}\frac{y_Q - y_P}{x_Q - x_P}ds_P - \pi\mu(Q).$$

If now we take M as B_δ, we obtain from (4)

$$w(B+) + w(B-) = -2\pi m + 2\int_C \mu(P)\frac{d}{ds_P}\tan^{-1}\frac{y_B - y_P}{x_B - x_P}ds_P,$$

$$w(B+) - w(B-) = -2\pi m + 2\pi\mu(B).$$

Taking M as A_δ,

$$w(A+) + w(A-) = -2\pi m + 2\int_C \mu(P)\frac{d}{ds_P}\tan^{-1}\frac{y_A - y_P}{x_A - x_P}ds_P,$$

$$w(A+) - w(A-) = -2\pi m.$$

We have, therefore,

$$\lim_{\delta\to 0+}\int_{A_\delta}^{B_\delta}\frac{dv}{dn}ds_\delta = w(A+) - w(B+)$$

$$= -\pi\mu(B) - \int_C \mu(P)\frac{d}{ds_P}H(A, B; P)ds_P,$$

where

$$H(A, B; P) = \tan^{-1}\frac{y_B - y_P}{x_B - x_P} - \tan^{-1}\frac{y_A - y_P}{x_A - x_P}.$$

Also

$$\lim_{\delta \to 0-} \int_{A_\delta}^{B_\delta} \frac{dv}{dn} ds_\delta = w(A-) - w(B-)$$

$$= \pi\mu(B) - \int_C \mu(P) \frac{d}{ds_P} H(A, B; P) ds_P.$$

Equation (3) may now be written in the form

$$\mu(B) = -\frac{\lambda}{\pi}[g(B) - g(A)] - \frac{\lambda}{\pi} \int_C \mu(P) \frac{d}{ds_P} H(A, B; P) ds_P.$$

Integrate by parts:

$$\mu(B) = -\frac{\lambda}{\pi}[g(B) - g(A)] - \frac{\lambda}{\pi}[\mu(P) H(A, B; P)]_C + \frac{\lambda}{\pi} \int_C H(A, B; P) d\mu(P)$$

$$= -\frac{\lambda}{\pi}[g(B) - g(A)] - m\lambda + \frac{\lambda}{\pi} \int_C H(A, B; P) d\mu(P).$$

Hence we have the following equation:

$$\mu(B) = G(B) + \frac{\lambda}{\pi} \int_C H(A, B; P) d\mu(P), \tag{5}$$

where $G(B) = -(\lambda/\pi)[g(B) - g(A)] - m\lambda$.

If $v(M)$ is a function of class (1) *which satisfies* (3), *the function $\mu(B)$, of limited variation and with regular discontinuities, satisfies the Stieltjes integral equation* (5). *Conversely, if $\mu(B)$ is a solution of* (5), *of limited variation and with regular discontinuities, the function $v(M)$ given by* (1) *will satisfy* (3).

4. **The equivalent classical equation.** We now proceed to show that this equation may be solved by means of one of classical form. Write

$$\mu(B) = G(B) + R(B). \tag{6}$$

We shall prove that if $\mu(B)$ is of limited variation, then $R(B)$ is absolutely continuous.

Substitute the value of $\mu(B)$ from (6) into (5):

$$G(B) + R(B) = G(B) + \frac{\lambda}{\pi} \int_C H(A, B; P) dG(P) + \frac{\lambda}{\pi} \int_C H(A, B; P) dR(P). \tag{7}$$

It suffices to consider one of the above integrals, say the first. We have

$$H(A,B;P) = \int_A^B \frac{d}{ds_Q} \tan^{-1} \frac{y_Q - y_P}{x_Q - x_P} ds_Q$$

$$= \int_A^B \frac{\cos(n_Q, QP)}{QP} ds_Q,$$

$$\int_C H(A,B;P)dG(P) = \int_C dG(P) \int_A^B \frac{\cos(n_Q, QP)}{QP} ds_Q.$$

We shall prove that we may invert the order of integration in the right hand member.*

Let $\{h_k(Q, P)\}$ denote a sequence of continuous functions such that

$$|h_k Q, P)| < \frac{|\cos(n_Q, QP)|}{QP} \qquad (k = 1,2, \cdots),$$

$$\lim_{k\to\infty} h_k(Q,P) = \frac{\cos(n_Q, QP)}{QP}.$$

We have

$$\int_C dG(P) \int_A^B h_k(Q,P)ds_Q = \int_A^B ds_Q \int_C h_k(Q,P)dG(P). \tag{8}$$

We shall show first that as $k\to\infty$ the right hand member of this equation approaches the limit

$$\int_A^B ds_Q \int_C \frac{\cos(n_Q, QP)}{QP} dG(P).$$

For this purpose we define a sequence $\{q_m(Q, P)\}$ as follows:

$$q_m(Q,P) = \frac{|\cos(n_Q, QP)|}{QP}, \text{ if } \frac{|\cos(n_Q, QP)|}{QP} < m,$$

$$= m, \quad \text{otherwise.}$$

The sequence

$$\int_C q_m(Q,P)dT(P),$$

* The proof is essentially that in Evans, The Rice Institute Pamphlet, vol. 7 (1920), p. 263, but is given here in more detail, for the sake of clearness.

where T is the total variation function of $G(P)$, is non-decreasing with m. Also

$$(8') \qquad \int_A^B ds_Q \int_C q_m(Q,P)dT(P) = \int_C dT(P) \int_A^B q_m(Q,P)ds_Q < \Gamma T(C)$$

and is therefore bounded for all m. Consequently

$$\lim_{m\to\infty} \int_C q_m(Q,P)dT(P) = \int_C \frac{|\cos(n_Q,QP)|}{QP} dT(P)$$

exists nearly everywhere on C, and is a summable function. Moreover, from (8'),

$$\int_A^B ds_Q \int_C \frac{|\cos(n_Q,QP)|}{QP} dT(P) = \int_C dT(P) \int_A^B \frac{|\cos(n_Q,QP)|}{QP} ds_Q,$$

by the fundamental properties of the Lebesgue and Daniell integrals.

Consequently, from the definition of $h_k(Q, P)$, we have

$$\lim_{k\to\infty} \int_C h_k(Q,P)dG(P) = \int_C \frac{\cos(n_Q,QP)}{QP} dG(P),$$

the integral on the right being a Daniell integral, which by (8') represents a summable function. Moreover

$$\left| \int_C h_k(Q,P)dG(P) \right| \leqq \int_C \frac{|\cos(n_Q,QP)|}{QP} dT(P),$$

a summable function. Hence, by the Lebesgue limit theorem,

$$\lim_{k\to\infty} \int_A^B ds_Q \int_C h_k(Q,P)dG(P) = \int_A^B ds_Q \int_C \frac{\cos(n_Q,QP)}{QP} dG(P).$$

As for the left hand side of (8),* we have

$$\lim_{k\to\infty} \int_A^B h_k(Q,P)ds_Q = \int_A^B \frac{\cos(n_Q,QP)}{QP} ds_Q.$$

In fact $\int_C dT(P) \int_A^B h_k(Q, P)ds_Q$ exists for all k. Moreover $\int_C dT(P) \cdot \int_A^B [|\cos(n_Q, QP)|/(QP)]ds_Q$ exists, since, as we have seen, the inside integral is a summable function. Finally

$$\int_A^B |h_k(Q,P)| ds_Q < \int_A^B \frac{|\cos(n_Q,QP)|}{QP} ds_Q.$$

* Evans, The Rice Institute Pamphlet, vol. 7 (1920), p. 257.

Consequently $\int_C dG(P)\int_A^B [\cos(n_Q, QP)/(QP)]ds_Q$ exists and

$$\lim_{k\to\infty} \int_C dG(P) \int_A^B h_k(Q,P)ds_Q = \int_C dG(P) \int_A^B \frac{\cos(n_Q,QP)}{QP} ds_Q.$$

We obtain therefore from (8), as $k\to\infty$,

$$\int_C dG(P) \int_A^B \frac{\cos(n_Q,QP)}{QP} ds_Q = \int_A^B ds_Q \int_C \frac{\cos(n_Q,QP)}{QP} dG(P).$$

The right hand side, being the indefinite integral of a summable function, is absolutely continuous. So therefore is the left hand side. Similarly for the remaining integral in (7). Hence $R(B)$ is absolutely continuous. If then we write (7) in the form

$$R(B) = \frac{\lambda}{\pi} \int_C H(A,B;P)[dG(P) + dR(P)],$$

the derivative nearly everywhere of $R(B)$ is

$$r(B) = \frac{\lambda}{\pi} \int_C \frac{\cos(n_B,BP)}{BP} [dG(P) + dR(P)].$$

Let $h(B) = (\lambda/\pi)\int_C [\cos(n_B, BP)/(BP)]dG(P)$. Then, since $R(B)$ is absolutely continuous:

$$R(B) = \int_A^B r(Q)ds_Q,$$

we have

$$r(B) = h(B) + \frac{\lambda}{\pi} \int_C \frac{\cos(n_B,BP)}{BP} r(P)ds_P. \tag{9}$$

Thus (9) is a consequence of (5). Conversely, if (9) has a summable solution $r(B)$, then $\mu(B)$ given by (6) is of limited variation and a solution of (5).

But (9) has a unique solution unless λ is one of a set of characteristic values, since (9) is the classical integral equation for the Neumann problem. Moreover it is known that $\lambda = -1$ is not a characteristic value for this equation.

We know that $\lambda = +1$ is a characteristic value, and that a necessary and sufficient condition that (9) have a solution is that

$$\int_C h(Q)ds_Q = 0. \tag{10}$$

But

$$\int_C h(Q)ds_Q = \frac{1}{\pi}\int_C ds_B \int_C \frac{\cos(n_B, BP)}{BP} dG(P).$$

By changing the order of integration,

$$\frac{1}{\pi}\int_C dG(P)\int_C \frac{\cos(n_B, BP)}{BP} ds_B = \frac{1}{\pi}\int_C \pi dG(P) = G(C) = 0.$$

Hence $g(C)=0$. Conversely, if $G(C)=0$, the analysis may be retraced, and we have (10). The condition $g(C)=0$ is therefore necessary and sufficient in order that (5) have a solution for $\lambda=+1$.

The function $v(M)$ determined by (1) is therefore that belonging to a particular solution of (5) by means of (1) plus the function determined by (1) where the $\mu(P)$ is a solution of the homogeneous equation corresponding to (5). But in this case

$$\mu(Q) = \int_A^Q \psi(P)ds_P,$$

where $\psi(P)$ is a solution of the homogeneous equation corresponding to (9). But we know the functions $v(M)=\int_C \log(1/(MP))\psi(P)ds_P$ to be constant. Theorem 2 is therefore proved.

In §§3, 4 we have assumed only that C is of class Γ, and have not required the existence of a curvature as in §§1, 2.*

The foregoing methods may be applied to obtain analogous theorems for potentials of a double layer, not only for the plane, but also for the three-dimensional case. The latter case, both for the single and the double layer, is to be discussed in a paper which the author is writing in conjunction with Professor Evans, who suggested the present investigation.

---

* If the curve has a curvature which is bounded for all points of the curve, then it is automatically of class Γ.

THE RICE INSTITUTE,
HOUSTON, TEXAS

ON A GENERALIZATION OF THE ASSOCIATIVE LAW*

BY

ANTON SUSCHKEWITSCH

1. In my investigations in group theory, I have observed that Lagrange's theorem (that the order of a group is divisible by the order of any subgroup) does not use for its proof the Associative Law in its whole extent; this law can be replaced by a more general postulate, "Postulate A", as I shall call it.

We shall represent our elements by capital italic letters; the operation upon them may be represented by a star $\star$, so that $A \star B$ signifies the result of this operation performed upon A and B. A set of elements closed under any operation $\star$ may be called "a group"; this word is thus used in a more general sense than is usual, since the operation $\star$ is arbitrary. The ordinary groups with a special well known operation may be called "classic" to distinguish them from our generalised groups. Sets and groups will be denoted by capital German letters.

POSTULATE A. In the equation

$$(X \star A) \star B = X \star C, \tag{1}$$

the element C depends upon the elements A and B only and not upon X. (We suppose here that X can be an arbitrary element of a finite group to which A, B and C belong also.)

The Associative Law is obviously a special case of this Postulate A, viz. if $C = A \star B$.

I have investigated the finite groups that are obtained by replacing the Associative Law in the system of postulates of Frobenius† by Postulate A. I have found the following properties of these groups.

I. Besides our operation $\star$ every group $\mathfrak{G}$ of our type has another operation that will be denoted by a little circle $\circ$ and defined as follows: the equation (1) being given, we write

$$C = A \circ B. \tag{2}$$

---

* Presented to the Society, October 27, 1928; received by the editors in July, 1927.

† Frobenius (*Über endliche Gruppen*, Berliner Sitzungsberichte, 1895) defines the classic finite groups by the four following postulates: 1. The operation that will be considered is uniform (eindeutig) and applicable to any two elements. 2. This operation is uniformly reversible (eindeutig umkehrbar), i.e. from $AB = AC$ or $BA = CA$ it follows that $B = C$. 3. The Associative Law is true for it. 4. The operation is "limited in its effect" (begrenzt in ihrer Wirkung); that signifies the possibility of forming finite groups of our elements.

It is easy to see, that $\mathfrak{G}$ is also a group relative to the operation $\circ$; we express this fact by writing $\mathfrak{G}(\circ)$; (analogously, $\mathfrak{G}(\star)$). I shall prove that $\mathfrak{G}(\circ)$ is classic.

II. The group $\mathfrak{G}(\star)$ has always a right unit (the same for all its elements).

III. If the group $\mathfrak{G}(\star)$ has also a single left unit for all its elements (that must necessarily coincide with the right unit), then the Associative Law is true for $\mathfrak{G}(\star)$; in this case $\mathfrak{G}(\star)$ is classic and the operations $\star$ and $\circ$ are identical.

It follows that in the systems of postulates of Moore* and Dickson† for the definition of classic groups the Associative Law can be replaced by Postulate A (or its left analogue).

IV. We associate with every element A of our group $\mathfrak{G}$ a substitution

$$\overline{A} \simeq \begin{pmatrix} X \\ X\star A \end{pmatrix}^{\ddagger}$$

whereby X runs over all elements of $\mathfrak{G}$. I prove that all those substitutions $\overline{A}$ (corresponding to each element A of $\mathfrak{G}$) form a substitution group $\overline{\mathfrak{G}}$ which is obviously classic and simply isomorphic with $\mathfrak{G}(\circ)$. Conversely, all such substitutions $\overline{A}$ form a group only if the Postulate A is true for $\mathfrak{G}(\star)$.

V. All groups of our type will be obtained from classic groups by making any substitution in the head-line of Cayley's table of a classic group. Moreover, it is sufficient to make only such substitutions as do not alter the unit of the classic group. Such a substitution may be denoted by α.

VI. $\mathfrak{H}(\star)$ being any subgroup of $\mathfrak{G}(\star)$, $\mathfrak{H}(\circ)$ is also a subgroup of $\mathfrak{G}(\circ)$, i.e. relative to the operation $\circ$. The converse is not true. Every subgroup $\mathfrak{H}$ of $\mathfrak{G}$ relative to $\circ$ is also a group relative to $\star$, if and only if the substitution α, which corresponds to $\mathfrak{G}(\star)$, has the following form:

$$\alpha = \begin{pmatrix} X \\ X^{l} \end{pmatrix},$$

the numbers l being relatively prime to the orders of corresponding elements X.

2. We shall prove now all the assertions of §1.

I. The group $\mathfrak{G}(\circ)$ is obviously uniformly reversible. Again:

$$[(X\star A)\star B]\star C = [X\star(A\circ B)]\star C = X\star[(A\circ B)\circ C];$$

---

* Moore, *A definition of abstract groups*, these Transactions, vol. 3 (1902).

† Dickson, *Definition of a group and a field by independent postulates*, these Transactions, vol. 6 (1905).

‡ The sign $\simeq$ signifies that we denote a complicated expression more simply with a single letter.

and on the other hand

$$[(X\star A)\star B]\star C = (X\star A)\star(B\circ C) = X\star[A\circ(B\circ C)] ;$$

and hence the Associative Law is true for $\mathfrak{G}$ $(\circ)$.

II. The classic group $\mathfrak{G}$ $(\circ)$ has always a unit E; it is such that

$$(X\star E)\star A = X\star(E\circ A) = X\star A ;$$

and therefore

$$X\star E = X \text{ for every } X ;$$

E is thus the right unit for $\mathfrak{G}$ $(\star)$.

III. Let E be a left unit of $\mathfrak{G}$ $(\star)$; we have, then,

$$(E\star A)\star B = E\star(A\star B) = A\star B ;$$

and hence by virtue of Postulate A for every element X

$$(X\star A)\star B = X\star(A\star B),$$

i.e. the Associative Law; hence $\mathfrak{G}$ $(\star)$ is classic, and $A\circ B = A\star B$.

IV. It follows from (1), by virtue of Postulate A, that $\overline{AB} = \overline{C}$; hence $\overline{\mathfrak{G}}$ is a substitution group simply isomorphic with $\mathfrak{G}$ $(\circ)$ (see (2)).

Conversely, let $\mathfrak{G}$ $(\star)$ be any finite uniformly reversible group and let $\overline{\mathfrak{G}}$ be the set of corresponding substitutions, which form also a (classic) group. Let

$$\overline{AB} = \overline{C}, \quad \text{or} \quad \begin{pmatrix} X \\ X\star A \end{pmatrix}\begin{pmatrix} X \\ X\star B \end{pmatrix} = \begin{pmatrix} X \\ X\star C \end{pmatrix};$$

since

$$\begin{pmatrix} X \\ X\star B \end{pmatrix} = \begin{pmatrix} X\star A \\ (X\star A)\star B \end{pmatrix},$$

it follows that

$$(X\star A)\star B = X\star C$$

for each element X of $\mathfrak{G}$ $(\star)$; hence Postulate A holds.

V. In the head-line of Cayley's table of $\mathfrak{G}$ $(\star)$ we make the following substitution:

$$\alpha \simeq \begin{pmatrix} X \\ E\star X \end{pmatrix}$$

(E being the right unit of $\mathfrak{G}$ $(\star)$). Let $E\star X = X'$. We define the third operation $\times$ as follows:

(3) $$A\star B = A \times B'.$$

The operation $\times$ is uniformly reversible and also associative; in fact we have from (1) and (3):

$$(X \times A') \times B' = X \times C', \tag{4}$$

C' depending on A' and B' only but not on X; let $X=E$; then $(E\times A')\times B'=E\times C'$; but we have $E\times X'=E\star X=X'$; hence $C'=A'\times B'$, and (4) gives us the Associative Law for $\times$; thus $\mathfrak{G}(\times)$ is classic. Again it follows from (2) that α gives an isomorphism between $\mathfrak{G}(\circ)$ and $\mathfrak{G}(\times)$.

Conversely, let $\mathfrak{G}(\times)$ be now a given classic group; we make in the headline of Cayley's table of $\mathfrak{G}(\times)$ any substitution

$$\beta \simeq \begin{pmatrix} X \\ \overline{X} \end{pmatrix}$$

and define a new operation $\star$ as follows:

$$A \times B = A \star \overline{B}.$$

The operation $\star$ is obviously uniform and uniformly reversible; the Postulate A is also true for $\star$; in fact, if

$$(X\star\overline{A})\star\overline{B} = X\star\overline{C},$$

we have

$$(X\star\overline{A})\star\overline{B} = (X \times A) \times B = X \times (A \times B)\,;$$

and

$$X\star\overline{C} = X \times C\,;$$

hence $A\times B=C$ and thus $\overline{C}$ depends upon $\overline{A}$ and $\overline{B}$ only.

E being the unit of $\mathfrak{G}(\times)$, $\overline{E}$ is the right unit for $\mathfrak{G}(\star)$; we have in fact $A\star\overline{E}=A\times E=A$.

I affirm that we can replace β by another substitution α, which does not alter E, and in this manner define a new operation, say $\square$, so that the group $\mathfrak{G}(\square)$ will be simply isomorphic with $\mathfrak{G}(\star)$ and have the right unit E. We take for α

$$\alpha \simeq \begin{pmatrix} X \\ \overline{X} \end{pmatrix}\begin{pmatrix} \overline{E}\star\overline{X} \\ \overline{X} \end{pmatrix}\begin{pmatrix} \overline{X} \\ X \end{pmatrix} = \begin{pmatrix} X \\ \overline{X} \end{pmatrix}\begin{pmatrix} \overline{E}\star\overline{X} \\ X \end{pmatrix} = \begin{pmatrix} X \\ \overline{X} \end{pmatrix}\begin{pmatrix} \overline{E}\times X \\ X \end{pmatrix};$$

let $\overline{E}\times X'=\overline{X}$; we can write then

$$\alpha = \begin{pmatrix} X \\ \overline{X} \end{pmatrix}\begin{pmatrix} \overline{X} \\ X' \end{pmatrix} = \begin{pmatrix} X \\ X' \end{pmatrix};$$

and so we define

$$A \times B = A \square B'.$$

We shall prove that the substitution

$$\alpha_1 \simeq \begin{pmatrix} \overline{X} \\ X' \end{pmatrix}$$

gives an isomorphism between the groups $\mathfrak{G}(\star)$ and $\mathfrak{G}(\square)$. Let

$$\overline{A} \star \overline{B} = \overline{C}\,; \tag{5}$$

we shall prove that we shall have also

$$A' \square B' = C'. \tag{6}$$

It follows from (5) that $\overline{A} \times B = \overline{C}$; but $\overline{A} = \overline{E} \times A'$, $\overline{C} = \overline{E} \times C'$; hence $(\overline{E} \times A') \times B = \overline{E} \times C'$; and since $\mathfrak{G}(\times)$ is classic,

$$\overline{E} \times (A' \times B) = \overline{E} \times C'\,;$$

hence $A' \times B = C'$, and so (6) is established.

VI. Let $\mathfrak{H} = P_1 + P_2 + P_3 + \cdots$,* $\mathfrak{H}(\star)$ being a subgroup of $\mathfrak{G}(\star)$. Let $(X \star P_\kappa) \star P_\lambda = X \star P_\mu$; the elements P_κ and P_λ of $\mathfrak{H}$ being given, the element P_μ exists also in $\mathfrak{H}$; by virtue of Postulate A we have $P_\kappa \circ P_\lambda = P_\mu$; hence $\mathfrak{H}(\circ)$ is also a group.

It follows, hence, that Lagrange's theorem is true for the groups $\mathfrak{G}(\star)$ of our type.

Let $\mathfrak{H}$ be now a subgroup of $\mathfrak{G}$ relative to $\circ$; we shall analyse the conditions by which $\mathfrak{H}$ is also a group relative to $\star$. Let α be the same substitution as in V, and

$$\mathfrak{H}' \simeq P_1' + P_2' + P_3' + \cdots.$$

($P_1', P_2', P_3', \cdots$ are elements in $\mathfrak{G}$ corresponding to $P_1, P_2, P_3, \cdots$, by virtue of α.) Since α gives an isomorphism between $\mathfrak{G}(\circ)$ and $\mathfrak{G}(\times)$ ($\times$ being the operation defined by (3)), $\mathfrak{H}'(\times)$ is also a group (relative to $\times$).

Let $\mathfrak{H}(\star)$ be also a group; then

$$P_\kappa \star P_\lambda = P_\kappa \times P_\lambda' \simeq P_\mu.$$

If P_λ' runs over all elements of $\mathfrak{H}'$, then P_μ runs over all elements of $\mathfrak{H}$, and conversely. Hence

$$P_\kappa \times \mathfrak{H}' = \mathfrak{H}$$

(for each P_κ of $\mathfrak{H}$). Consequently $\mathfrak{H}$ is one of the partitions of $\mathfrak{G}(\times)$ relative to $\mathfrak{H}'(\times)$†. This condition is obviously also sufficient for $\mathfrak{H}(\star)$ to be a group.

* The sign + signifies that the elements $P_1, P_2, \cdots$ form a set $\mathfrak{H}$.

† Hilton, *An Introduction to the Theory of Groups of Finite Order*, Oxford, 1908, p. 58.

Since the substitution α does not alter the unit E of $\mathfrak{G}(\times)$, $\mathfrak{H}$ and $\mathfrak{H}'$ must be identically equal to each other, because both of them have a common element E.

We shall now analyse the conditions by which *every* subgroup $\mathfrak{H}(o)$ of $\mathfrak{G}(o)$ is also a group relative to $\star$. Then we must have $\mathfrak{H}'=\mathfrak{H}$ (our notation remains as above) for every subgroup $\mathfrak{H}(o)$. We take $\mathfrak{H}'=\mathfrak{H}(\times)=\{P\}$, a cyclic group, P being an arbitrary element of $\mathfrak{G}$. Since $\{P\}$ must be also a group relative to $\star$, we have

$$P^k \star P = P^k \times P^l ; \tag{7}$$

consequently for each element X of $\mathfrak{G}$ also,

$$X \star P = X \times P^l. \tag{7'}$$

More generally,

$$X \star P^\kappa = X \times P^\lambda. \tag{8}$$

To every exponent κ in (8) there corresponds one and only one exponent λ and vice versa. This must be true for each element P of $\mathfrak{G}$; if we take P^k instead of P, we obtain, in the same manner as in (8),

$$X \star P^{k\mu} = X \times P^{k\nu} ; \tag{9}$$

for every μ there is a definite ν and vice versa. Let m be the order of P, and d the greatest common divisor of k and m; then m/d is the order of P^k and each exponent $k\mu$ and $k\nu$ in (9) is divisible by d. Conversely, if one of the exponents κ, λ in (8) is prime to m, the other is also prime to m. Consequently the exponent l in (7) or (7') must be prime to m. Thus α has in this case the following form:

$$\alpha = \begin{pmatrix} X \\ X^l \end{pmatrix},$$

where the numbers l are prime to the orders of corresponding elements X. This condition is not only necessary but also sufficient: if it holds, then every cyclic subgroup $\{P\}$ of $\mathfrak{G}(\times)$ is also a group relative to $\star$. But hence *every* subgroup $\mathfrak{H}(\times)$ of $\mathfrak{G}(\times)$ is also a group relative to $\star$. Q and P being any two elements of $\mathfrak{H}$, we have in fact $Q \star P = Q \times P^l$; thus $Q \star P$ belongs also to $\mathfrak{H}$.

We can take, in particular, a substitution α of the following form:

$$\alpha = \begin{pmatrix} X \\ X^\tau \end{pmatrix},$$

where τ is the same for each element X and relatively prime to the order of our group $\mathfrak{G}$.

3. We shall consider now a special case of Postulate A, that is, however, more general than the Associative Law.

POSTULATE B. In the equation

$$(X\star A)\star B = X\star(A\star B_1), \tag{10}$$

the elements B and B_1 depend only upon each other; every B is completely defined by the corresponding B_1, and conversely.

This postulate can be expressed in another form as follows:

POSTULATE B′. If

$$A\star B = C\star D \tag{11}$$

and if K is an arbitrary element, then

$$A\star(B\star K) = C\star(D\star K). \tag{12}$$

We prove first that Postulate B′ follows from Postulate B. Let R be an element such that

$$A\star(B\star K) = (A\star B)\star R\ ; \tag{13}$$

R depends upon K only (by Postulate B). Again it follows from (11) that

$$(A\star B)\star R = (C\star D)\star R\ ; \tag{14}$$

and by Postulate B it follows from (13) that

$$C\star(D\star K) = (C\star D)\star R\ ; \tag{15}$$

hence, from (11), (13), (14), (15) it follows that (12) holds.

Second, we prove that Postulate B follows also from the Postulate B′. For that purpose we shall prove the following lemma:

LEMMA: *If Postulate* B′ *is true for a (uniformly reversible) group* $\mathfrak{G}(\star)$, *there exists in* $\mathfrak{G}(\star)$ *a right unit (for all elements of* $\mathfrak{G}(\star)$).

If B is a given element, there always exists in $\mathfrak{G}(\star)$ an element E such that

$$B\star E = B. \tag{16}$$

Let D be an arbitrary element of $\mathfrak{G}(\star)$; if A is also a given element, there always exists an element C, for which

$$A\star B = C\star D\ ; \tag{11}$$

by virtue of Postulate B′ we have, then,

$$A\star(B\star E) = C\star(D\star E)\ ; \tag{17}$$

and by virtue of (16) and (11) it follows from (17) that E is the right unit for *every* element D.

Assume now that $A\star B=C=C\star E$; in the hypothesis of Postulate B′ we have, K being an arbitrary element,

$$A\star(B\star K) = C\star(E\star K) = (A\star B)\star(E\star K) ;$$

and thus we have in (13) $E\star K=R$; this shows that Postulate B holds for our group.

Since the groups with Postulate B form a special case of groups with Postulate A, they can be obtained in the same manner as groups with Postulate A (§1, V). We must now examine what must be the substitution α (§1, V), in order that we may obtain a group $\mathfrak{G}(\star)$ with Postulate B. The answer is given by the following theorem:

THEOREM. *If the group $\mathfrak{G}(\star)$ is obtained from the classic group $\mathfrak{G}(\times)$ by means of the substitution α, Postulate* B *is true for $\mathfrak{G}(\star)$ if and only if α is an automorphism of the group $\mathfrak{G}(\times)$. In this case α is also an automorphism for $\mathfrak{G}(\star)$, and the operations $\circ$ and $\times$ coincide with each other.*

Let $\mathfrak{G}(\star)$ be a group with Postulate B. The equation (10) gives a dependence of B and B_1 upon each other; this dependence is given by a substitution, that we denote symbolically by $\binom{X}{X_1}$. Let $A=E$ (the right unit of $\mathfrak{G}(\star)$) in (10); then

$$X\star B = X\star(E\star B_1) ;$$

and hence

$$B = E\star B_1. \tag{18}$$

Let

$$\alpha \simeq \binom{X}{X'} ;$$

we have then

$$B = E\star B_1 = E \times (B_1)' = (B_1)' ;$$

and hence

$$\alpha = \binom{X_1}{X} = \binom{X}{X_1}^{-1}.$$

Moreover it follows from (10), if we use the notation $A\star B_1\simeq C$, that

$$C = A\circ B = A\star B_1 = A \times (B_1)' = A \times B ;$$

thus the operations $\circ$ and $\times$ coincide.

Conversely, suppose that the operations $\circ$ and $\times$ coincide. Let

$$\alpha = \binom{X}{X'} = \binom{X_1}{X};$$

we have then

$$X \star A = X \times A' ; \tag{19}$$

and

$$(X \star A) \star B = X \star (A \times B) = X \star (A \star B_1),$$

and that is Postulate B, because B and B_1 depend only on each other. Again, by (19),

$$(X \star A) \star B = X \star C = (X \times A') \times B' = X \times C' ;$$

and hence

$$A \times B = C, \quad A' \times B' = C' ;$$

this shows us that α is an automorphism of $\mathfrak{G}(\times)$.

Conversely, let α be an automorphism of $\mathfrak{G}(\times)$; then

$$(X \star A) \star B = (X \times A') \times B' = X \times (A' \times B')$$
$$= X \times (A \times B)' = X \star (A \times B) = X \star (A \star B_1) ;$$

and thus Postulate B holds.

It remains to prove that α is in this case an automorphism of $\mathfrak{G}(\star)$ also. We have in fact

$$(A \star B)' = (A \times B')' = A' \times (B')' = A' \star B'.$$

4. In the theory of uniformly reversible groups we can consider the operations inverse to the operation of a given group. Since the operation of our group is performed upon *two* elements (viz. $X \star Y$), two inverse operations exist according as the left or the right of these two elements is unknown to us.

If the commutative law is true for our group, such a group has only *one* inverse operation and only one "inverse group" (i.e. the group relative to the inverse operation). But although a general classic group has two "inverse groups," it has only one inverse *operation* (abstractly considered), because the properties of the operation of a classic group are "symmetric," i.e. the same on both sides; two "inverse groups" of a classic group are simply isomorphic to each other (if our notations are conveniently chosen); this follows from the fact that a classic group is always "anti-isomorphic" to itself, i.e. there always exists such a substitution $\binom{X}{\overline{X}}$ of elements of a classic group, that if A, B correspond respectively to $\overline{A}$, $\overline{B}$, then AB corresponds to $\overline{B}\,\overline{A}$; we can take, for example, $\overline{X} = X^{-1}$.

The operation of a finite classic group $\mathfrak{G}$ may be denoted by $\times$, the two inverse operations by $\triangle$ and $\triangledown$; more precisely,

$$\text{if } A \times B = C, \quad \text{then } C \triangle B = A, \quad C \triangledown A = B.$$

Both inverse groups $\mathfrak{G}(\triangle)$, $\mathfrak{G}(\triangledown)$ are finite and uniformly reversible but not associative. Let us consider what influence the associative law of the operation $\times$ makes on the operations $\triangle$ and $\triangledown$. Let $(A \times B) \times C = A \times (B \times C) \simeq R$; $A \times B \simeq P$; $B \times C \simeq Q$; then $P \times C = A \times Q = R$. Hence $P \triangle B = A$, $Q \triangle C = B$, $R \triangle C = P$, $R \triangle Q = A$; consequently

$$(R \triangle C) \triangle B = R \triangle Q, \quad \text{and} \quad B \times C = Q.$$

[Or $P \triangledown A = B$, $Q \triangledown B = C$, $R \triangledown P = C$, $R \triangledown A = Q$, $(R \triangledown A) \triangledown B = R \triangledown P$, and $P = A \times B$.] This is Postulate A, that is true for the operation $\triangle$ (and for $\triangledown$). But the operations $\triangle$ and $\triangledown$ are subject to still another postulate, viz.:

POSTULATE J. Every element X satisfies the equation

$$X \triangle X = E \quad (\text{or} \quad X \triangledown X = E),$$

where E is a determined element (the unit of the direct operation $\times$).

THEOREM 1. *A finite uniformly reversible group $\mathfrak{G}(\star)$ is an "inverse" to a classic group, if and only if it is subject to the postulates* A *and* J.

Only one part of this theorem remains for us to prove. Let $\mathfrak{G}(\star)$ be subject to the postulates A and J. We use the same notation as before; if $A \star B = C$, then $C \triangle B = A$. We must prove that $\mathfrak{G}(\triangle)$ is classic. Obviously the operation $\triangle$ is uniform and uniformly reversible. Again, we have $(X \star A) \star B = X \star C \simeq Z$; C depends upon A and B only; let $X \star A \simeq Y$; then $Y \star B = X \star C = Z$; $Z \triangle C = X$, $Z \triangle B = Y$, $Y \triangle A = X$; thus

$$(Z \triangle B) \triangle A = Z \triangle C,$$

which is Postulate A for the operation $\triangle$. It follows from Postulate J, that the group $\mathfrak{G}(\triangle)$ has a left unit E; and hence (see §1, III) $\mathfrak{G}(\triangle)$ is classic.

We consider a special case, when our classic group is abelian. We obtain then

THEOREM 2. *A finite uniformly reversible group $\mathfrak{G}(\star)$ is an "inverse" to an abelian group, if and only if it is subject to the postulates* B *and* J.

Let $\mathfrak{G}(\star)$ be subject to the postulates B and J; by the preceding theorem the inverse group $\mathfrak{G}(\triangle)$ is classic; it remains for us to show that $\mathfrak{G}(\triangle)$ is commutative. We have

$$(X \star A) \star B = X \star (A \star B_1), \tag{10}$$

B and B_1 depending only upon each other. Let $A \star B_1 = C$; then (as in the preceding theorem) $B \triangle A = C$, $C \triangle B_1 = A$; hence

$$B \triangle A \triangle B_1 = A. \tag{20}$$

We write this without brackets, because the Associative Law is true for $\triangle$; (20) is true for each element A; we take $A = E$ (unit); then $B \triangle B_1 = E$; $B_1 = B^{-1}$; and thus from (20) it follows that $A \triangle B = B \triangle A$; i. e., the Commutative Law holds for $\triangle$.

Conversely, let $\mathfrak{G}$ $(\triangle)$ be an abelian group; we must prove that B and B_1 in (10) depend only upon each other. But (20) gives $A \star B^{-1} = B \triangle A = C$; hence $B_1 = B^{-1}$ in (10), and Postulate B holds for $\star$.

The postulates B and J are characteristic for the operation of division. Thus it is possible, for example, to construct an abstract theory of proportions.

Supplement

Example I. A group with Postulate A but not classic (see Table 1). This group is obtained from the symmetric group of 6th order by making in the head-line of Cayley's table of this group (see Table 2) the following substitution:

$$\begin{pmatrix} E & A & B & C & D & F \\ E & C & D & A & F & B \end{pmatrix}.$$

Example II. A group with Postulate B but not classic (see Table 3). This group is obtained from the same symmetric group by making in the head-line of Table 2 the following substitution:

$$\begin{pmatrix} E & A & B & C & D & F \\ E & B & A & C & F & D \end{pmatrix}.$$

This substitution gives an automorphism of the symmetric group of 6th order.

	E	A	B	C	D	F
E	E	C	F	A	B	D
A	A	D	B	E	F	C
B	B	F	C	D	E	A
C	C	E	A	F	D	B
D	D	A	E	B	C	F
F	F	B	D	C	A	E

Table 1

	E	A	B	C	D	F
E	E	A	B	C	D	F
A	A	E	F	D	C	B
B	B	D	E	F	A	C
C	C	F	D	E	B	A
D	D	B	C	A	F	E
F	F	C	A	B	E	D

Table 2

	E	A	B	C	D	F
E	E	B	A	C	F	D
A	A	F	E	D	B	C
B	B	E	D	F	C	A
C	C	D	F	E	A	B
D	D	C	B	A	E	F
F	F	A	C	B	D	E

Table 3

Voronesh University,
Voronesh, Russia

A NOTE ON CLOSEST APPROXIMATION*

BY

DUNHAM JACKSON

Let $f(x)$ be a given continuous function of period 2π, and $T_n(x)$ a trigonometric sum of the nth order, ultimately to be chosen so as to give an approximate representation of $f(x)$, but for the present arbitrary. A numerical measure of the discrepancy between $f(x)$ and $T_n(x)$ can be defined in an infinite variety of ways. One of the most important of such measures is the maximum of the absolute value of $f(x)-T_n(x)$; another is the integral of the square of this difference, extended over an interval of length 2π. If $T_n(x)$ is chosen so as to minimize the latter quantity, it is the partial sum of the Fourier series for $f(x)$. If it minimizes the former, it is the approximating sum in the sense of Tchebychef.

This note is concerned with a measure of discrepancy which is in a certain sense qualitatively intermediate between the two preceding ones, representing a transition from one to the other. It is defined by forming the integral of the square of the error over an interval of length h, where h is a given positive number, in general different from 2π, and then taking the maximum of this integral, regarded as a function of the position of the initial point of the interval.

Let the maximum of the integral, for given $T_n(x)$ and given h, be denoted by I_{nh}. The particular approximating function to be studied is the sum $T_{nh}(x)$ characterized among all trigonometric sums of the nth order as the one which makes I_{nh} a minimum. It reduces to the Fourier sum for $h=2\pi$, the value of the integral being then independent of the situation of the interval, and it approaches the Tchebychef sum (as will be shown presently) when h approaches zero. In view of the fact that (for $0<h<2\pi$) it occupies this intermediate position with respect to well known approximating sums, it is not surprising that $T_{nh}(x)$ converges uniformly toward $f(x)$, under fairly general hypotheses, if h is held fast and n is allowed to become infinite. The main purpose of the paper is to prove that this is the case, and incidentally to point out that the method is susceptible of much more general application. (As far as the validity of the reasoning is concerned, values of h greater than 2π, as well as those less than 2π, may be admitted from the beginning.)

---

* Presented to the Society at Chicago, April 7, 1928; received by the editors in August, 1928.

BOSTON UNIVERSITY
COLLEGE OF LIBERAL ARTS
LIBRARY

The existence of a sum $T_{n\lambda}(x)$ with the required minimizing property follows almost immediately from well known existence proofs for the case in which the integral is extended over a period. As the whole problem is trivial when $f(x)$ is identically a trigonometric sum of the nth order, it will be understood that this case is ruled out. It is readily seen that $I_{n\lambda}$ is a continuous function of the coefficients in $T_n(x)$; and then the essential thing is to show that the assignment of an upper bound for $I_{n\lambda}$ prescribes at the same time an upper bound for the magnitude of the coefficients, so that the range within which the minimizing coefficients are to be sought may be regarded as closed. To say that an upper bound is fixed for the size of the coefficients is the same thing as saying that an upper bound is prescribed for the maximum of the absolute value of $T_n(x)$ itself.* The existence of the desired upper bound for $|T_n(x)|$ is known† when $h=2\pi$. More definitely, if

$$\int_0^{2\pi} [f(x) - T_n(x)]^2 dx \leqq G,$$

then

$$\max |T_n(x)| \leqq QG^{1/2}, \tag{1}$$

where Q is independent of the coefficients (being fixed when $f(x)$ and n are given).‡ If $h>2\pi$, and if

$$\int_a^{a+h} [f(x) - T_n(x)]^2 dx \leqq G,$$

the integral over a period satisfies the same inequality, and (1) follows at once. When $h<2\pi$, let p be an integer such that $ph \geqq 2\pi$. If $I_{n\lambda} \leqq G$, the integral of the square of the error over any interval of length less than or equal to h can not exceed G; a period can be divided into p such intervals, and consequently

$$\int_0^{2\pi} [f(x) - T_n(x)]^2 dx \leqq pG,$$

whence $\max |T_n(x)| \leqq Q(pG)^{1/2}$, a relation which serves as well as (1) for

* The equivalence of the two statements is well known; for a generalization due to Sibirani, see for example the passage cited in the next footnote.

† For a formulation covering the case now in question and others to be referred to later, see for example D. Jackson, *A generalized problem in weighted approximation*, these Transactions, vol. 26 (1924), pp. 133–154; pp. 133–139.

‡ This form of statement is not taken directly from the passage cited, but is readily deduced from it.

the purpose in hand. So the existence of the minimizing $T_{n\lambda}(x)$ is assured.

The uniqueness of the minimizing sum can be proved by an argument already used in connection with numerous other problems. Let $\gamma_{n\lambda}$ denote the minimum of $I_{n\lambda}$. Suppose there were two distinct sums $T_{n\lambda 1}(x)$ and $T_{n\lambda 2}(x)$, for each of which the corresponding $I_{n\lambda}$ is equal to $\gamma_{n\lambda}$. Let

$$T_{n\lambda 3}(x) = \tfrac{1}{2}(T_{n\lambda 1} + T_{n\lambda 2}).$$

Then

$$f(x) - T_{n\lambda 3}(x) = \tfrac{1}{2}\{[f(x) - T_{n\lambda 1}(x)] + [f(x) - T_{n\lambda 2}(x)]\},$$

and

$$[f(x) - T_{n\lambda 3}(x)]^2 \leqq \tfrac{1}{2}\{[f(x) - T_{n\lambda 1}(x)]^2 + [f(x) - T_{n\lambda 2}(x)]^2\},$$

the equality being ruled out wherever $T_{n\lambda 1} \neq T_{n\lambda 2}$, and being possible therefore only for a finite number of values of x at most in any finite interval. For any specified interval of length h, the integral of either of the squares on the right can not exceed the corresponding $I_{n\lambda}$, and so can not exceed $\gamma_{n\lambda}$; for any interval $(a, a+h)$, therefore, by virtue of the fact that the inequality holds almost everywhere,

$$\int_a^{a+h} [f(x) - T_{n\lambda 3}(x)]^2 dx < \gamma_{n\lambda},$$

which contradicts the definition of $\gamma_{n\lambda}$.

It may be pointed out next that for given n the sums $T_{n\lambda}(x)$ are uniformly bounded for all values of h. Let M be the maximum of $|f(x)|$, and let $M_{n\lambda}$ be the maximum of $|T_{n\lambda}(x)|$. From its definition, the quantity $\gamma_{n\lambda}$, the maximum of the integral of $[f(x) - T_{n\lambda}(x)]^2$ over an interval of length h, can not exceed the corresponding maximum $I_{n\lambda}$ formed for the particular trigonometric sum $T_n(x) \equiv 0$. But this $I_{n\lambda}$ is merely the integral of $[f(x)]^2$ over some interval of length h, and can not exceed M^2h. So

$$\gamma_{n\lambda} \leqq M^2 h \tag{2}$$

for all values of h.

In the relatively trivial case that $n = 0$, $T_{n\lambda}(x) =$ constant, it is seen at once that $|T_{n\lambda}| \leqq 2M$. For if $|T_{n\lambda}|$ were greater than $2M$, $|f(x) - T_{n\lambda}|$ would be everywhere greater than M, and the integral of its square over any interval of length h would be greater than M^2h, in violation of (2).

To proceed to the calculation of an upper bound for $M_{n\lambda}$ when $n > 0$, suppose first that $0 < h \leqq 1/n$. In this case it is to be shown that $M_{n\lambda} \leqq 4M$. Let x_0 be a value such that $|T_{n\lambda}(x_0)| = M_{n\lambda}$. By Bernstein's theorem, $|T'_{n\lambda}(x)| \leqq nM_{n\lambda}$, and for $|x - x_0| \leqq 1/(2n)$,

(3) $\quad |T_{nh}(x) - T_{nh}(x_0)| \leqq nM_{nh}|x - x_0| \leqq \frac{1}{2}M_{nh}, \quad |T_{nh}(x)| \geqq \frac{1}{2}M_{nh}.$

Let it be assumed, as a supposition to be disproved, that $M_{nh} > 4M$. Then (3) means that $|T_{nh}(x)| > 2M$ throughout an interval of length $1/n$, and so throughout an interval of length h, and in this interval $|f(x) - T_{nh}(x)| > M$. But this implies that $\gamma_{nh} > M^2h$, in contradiction with (2). So M_{nh} has the upper bound $4M$, independent of h, for the indicated range of values of h.

Secondly, suppose $1/n \leqq h \leqq 2\pi$. From (2), $\gamma_{nh} \leqq M^2h \leqq 2\pi M^2$. Let p be an integer greater than $2\pi n$, for definiteness the smallest such integer. Then the interval $(0, 2\pi)$ can be subdivided into p intervals, each of length less than $1/n$, and so less than h. But the integral of $[f(x) - T_{nh}(x)]^2$ over any such interval can not exceed γ_{nh}, and therefore

$$\int_0^{2\pi} [f(x) - T_{nh}(x)]^2 dx \leqq p\gamma_{nh} \leqq 2p\pi M^2.$$

Consequently

$$M_{nh} \leqq (2p\pi)^{1/2} QM,$$

where Q has the same meaning as in (1).

Let it be supposed finally that $2k\pi \leqq h \leqq 2(k+1)\pi$, where k is a positive integer. Then

$$\int_0^{2k\pi} [f(x) - T_{nh}(x)]^2 dx \leqq \int_0^{h} [f(x) - T_{nh}(x)]^2 dx \leqq \gamma_{nh}$$
$$\leqq M^2h \leqq 2(k+1)\pi M^2,$$

and

$$\int_0^{2\pi} [f(x) - T_{nh}(x)]^2 dx = (1/k) \int_0^{2k\pi} [f(x) - T_{nh}(x)]^2 dx$$
$$\leqq 2(k+1)\pi M^2/k \leqq 4\pi M^2,$$

so that

$$M_{nh} \leqq (4\pi)^{1/2} QM.$$

If M_0 is defined as the largest of the quantities $4M$, $(2p\pi)^{1/2}QM$, $(4\pi)^{1/2}QM$ (in other words, the larger of the first two of these quantities, since the third is evidently less than the second), this M_0 is independent of h, and $M_{nh} \leqq M_0$ throughout.

It is possible now to justify the assertion made in one of the introductory paragraphs, that $T_{nh}(x)$ approaches the Tchebychef sum when h approaches zero. Let the Tchebychef sum, reducing the maximum of the error to the smallest possible value, be denoted by $Z_n(x)$. Let ϵ be an arbitrary positive quantity. For the moment, let $U_n(x)$ be a general notation for trigonometric sums of the nth order subject to the restriction that at least one coefficient

differs from the corresponding coefficient in $Z_n(x)$ by not less than ϵ. Let $Z_{n1}(x)$ be chosen among all sums $U_n(x)$ as one for which the maximum of $|f(x)-Z_{n1}(x)|$ has the smallest possible value. It is clear from the considerations leading to the proof of existence of the Tchebychef sum that there is a $Z_{n1}(x)$ having the required minimizing property; it is immaterial for the present argument whether the minimum is unique or not. Let L be the maximum of $|f(x)-Z_n(x)|$, and $L_1=L+\eta$ the maximum of $|f(x)-Z_{n1}(x)|$. Since the Tchebychef sum at any rate is unique, it is certain that $\eta>0$. Being continuous and periodic, $f(x)$ is uniformly continuous; let δ_1 be a positive number such that $|f(x')-f(x'')|\leqq\frac{1}{4}\eta$ whenever $|x'-x''|\leqq\delta_1$. Let δ be the smaller of the numbers δ_1 and $\eta/(4nM_0)$, where M_0 is defined as above. It will be shown that if $h<\delta$, $T_{nh}(x)$ can not belong to the class of sums $U_n(x)$; that is, to every $\epsilon>0$ it is possible to assign a $\delta>0$ such that if $h<\delta$, no coefficient in $T_{nh}(x)$ can differ from the corresponding coefficient in $Z_n(x)$ by so much as ϵ.

For the sake of arriving at a contradiction, suppose $T_{nh}(x)$ is of the form $U_n(x)$, while $h<\delta$. There must be a point x_0 such that $|f(x_0)-T_{nh}(x_0)|\geqq L+\eta$. For $|x-x_0|\leqq\delta$, $|f(x)-f(x_0)|\leqq\frac{1}{4}\eta$. On the other hand, $|T_{nh}(x)|\leqq M_0$, and $|T'_{nh}(x)|\leqq nM_0$. For $|x-x_0|\leqq\delta$, therefore,

$$|T_{nh}(x)-T_{nh}(x_0)|\leqq nM_0\delta\leqq\tfrac{1}{4}\eta,$$

whence it follows further that

$$|[f(x)-f(x_0)]-[T_{nh}(x)-T_{nh}(x_0)]|\leqq\tfrac{1}{2}\eta,$$
$$|f(x)-T_{nh}(x)|\geqq L+\tfrac{1}{2}\eta.$$

The last relation holds throughout an interval of length δ, and so throughout an interval of length h. Consequently γ_{nh}, the maximum of the integral of $[f(x)-T_{nh}(x)]^2$ over an interval of length h, is at least $h(L+\frac{1}{2}\eta)^2$. But if the corresponding maximum I_{nh} is formed for the sum $T_n(x)\equiv Z_n(x)$, the error never exceeds L, and the integral I_{nh} can not exceed hL^2. This contradicts the definition of γ_{nh} as the smallest value of I_{nh}. So the facts are as originally stated.

The proof of convergence for $n=\infty$ is analogous to others which the writer has given in recent papers. Let $t_n(x)$ be an arbitrary trigonometric sum of the nth order, and ϵ_n the maximum of $|f(x)-t_n(x)|$. Let

$$f(x)-t_n(x)=\tau_n(x),\qquad T_{nh}(x)-t_n(x)=\tau_n(x);$$

it is to be understood that h is held fast throughout the present demonstration. Let μ_n be the maximum of $|\tau_n(x)|$, and x_0 a point such that $|\tau_n(x_0)|$

$=\mu_n$. According to Bernstein's theorem, $|\tau_n'(x)| \leqq n\mu_n$, and, for values of x in the interval $|x-x_0| \leqq 1/(2n)$,

$$|\tau_n(x) - \tau_n(x_0)| \leqq n\mu_n |x - x_0| \leqq \tfrac{1}{2}\mu_n, \quad |\tau_n(x)| \geqq \tfrac{1}{2}\mu_n.$$

As one of two conceivable alternatives, suppose $\mu_n \geqq 4\epsilon_n$. Then $|r_n(x)| \leqq \epsilon_n \leqq \frac{1}{4}\mu_n$, and

$$|r_n(x) - \tau_n(x)| \geqq \tfrac{1}{4}\mu_n$$

throughout the interval specified. The length of this interval is $1/n$, *which is less than* h, *as soon as* n *is sufficiently large.* The integral of $[r_n(x)-\tau_n(x)]^2$ over the interval is at least $(\frac{1}{4}\mu_n)^2/n$. But from the definitions of the quantities involved, $r_n(x)-\tau_n(x)$ is identical with $f(x)-T_{nh}(x)$, and the integral of $[f(x)-T_{nh}(x)]^2$ over any interval of length $\leqq h$ can not exceed γ_{nh}. Consequently

$$(\tfrac{1}{4}\mu_n)^2/n \leqq \gamma_{nh}, \qquad \mu_n \leqq 4(n\gamma_{nh})^{1/2}.$$

The other alternative is to suppose directly that $\mu_n < 4\epsilon_n$. In either case,

$$\mu_n \leqq 4(n\gamma_{nh})^{1/2} + 4\epsilon_n,$$

for all values of n from a certain point on. Since μ_n is the maximum of $|\tau_n(x)|$, and since $|r_n(x)| \leqq \epsilon_n$,

$$|f(x) - T_{nh}(x)| \equiv |r_n(x) - \tau_n(x)| \leqq 4(n\gamma_{nh})^{1/2} + 5\epsilon_n.$$

The function $f(x)$ being continuous, it is certainly possible, by Weierstrass's theorem, to choose sums $t_n(x)$ for the successive values of n so that $\lim_{n=\infty} \epsilon_n = 0$. *For uniform convergence of* $T_{nh}(x)$ *toward* $f(x)$, *therefore, it is sufficient that* $\lim_{n=\infty} n\gamma_{nh} = 0$.

Let the modulus of continuity of $f(x)$ be denoted by $\omega(\delta)$. It is known* that for all positive integral values of n trigonometric sums of the nth order can be constructed so as to differ from $f(x)$ by less than a constant multiple of $\omega(2\pi/n)$ for all values of x. If the corresponding quantity I_{nh} is calculated in each case, it does not exceed a constant multiple of $[\omega(2\pi/n)]^2$. Hence

$$\gamma_{nh} \leqq C[\omega(2\pi/n)]^2,$$

where C is independent of n. *The sums* $T_{nh}(x)$ *will converge uniformly toward* $f(x)$, *if* $f(x)$ *has a modulus of continuity* $\omega(\delta)$ *such that* $\lim_{\delta=0} \omega(\delta)/\delta^{1/2} = 0$.

---

* See for example D. Jackson, *On the approximate representation of an indefinite integral* ... these Transactions, vol. 14 (1913), pp. 343–364; p. 350.

Some of the generalizations that come to mind may be mentioned though without any attempt at exhaustiveness. The generalizations are concerned for the most part with the existence and uniqueness of an approximating function for given n, and its convergence toward $f(x)$ as n becomes infinite.

As a matter of course, the square of the error may be replaced by the mth power of its absolute value, in the integrals leading to the definition of the approximating sums. In the proof of uniqueness it is to be assumed, for simplicity at least, that $m>1$; in the convergence theorem (valid for $m>0$) the hypothesis on $f(x)$ is to be appropriately modified. While the direct connection with the Fourier sum lapses, it is still possible in this case to show that the approximating function approaches the Tchebychef sum as a limit when h approaches zero.

A weight function may be introduced in the integrals, the assignment of weights being regarded as fixed either with respect to the origin from which x is measured, or with respect to the initial point of the interval of integration; that is, I_{nh} may be defined as the maximum of

$$\int_a^{a+h} \rho(x)[f(x) - T_n(x)]^2 dx$$

or of

$$\int_a^{a+h} \rho(x-a)[f(x) - T_n(x)]^2 dx$$

as a varies, ρ being a given bounded and measurable function of its argument with a positive lower bound.

The maximum of the integral over a single interval of length h may be replaced by the maximum of the sum of the integrals over N intervals of given aggregate length $h<2\pi$, where N is a given integer, and the intervals are supposed non-overlapping and contained within a single period. (Intervals having a common end point are admitted, and accordingly a set of fewer than N intervals is to be regarded for the purposes of the definition as a special case of a set of N intervals, since one interval can be subdivided to make up the requisite number.)

As an alternative, the class of all intervals of length h, among which a maximum value for the integral is to be sought in defining I_{nh}, may be replaced by the class of all point sets congruent to a given measurable set, the assumption being made, for simplicity at any rate, that this set contains at least one interval.

Finally, let $h_1, h_2, \cdots, h_N$ and $c_1, c_2, \cdots, c_N$ be two given sets of N

positive numbers each; and let the quantities I_{nh} calculated in accordance with the original definition for intervals of lengths $h_1, \cdots, h_N$ respectively be denoted by $I_{n1}, \cdots, I_{nN}$. Then an approximating sum $T_n(x)$ may be determined so as to minimize the combination $c_1I_{n1}+c_2I_{n2}+\cdots+c_NI_{nN}$. (This is suggested as a generalization of the original definition for $h>2\pi$; for example, if $h_1=2\pi$, $h_2<2\pi$, $c_1=c_2=1$, the new definition is equivalent to the old one as formulated for a single interval of length $2\pi+h_2$.)

In all these cases, proofs of existence, uniqueness, and convergence can be carried through without any considerable modification of the treatment given in the text.

UNIVERSITY OF MINNESOTA,
MINNEAPOLIS, MINN.

HOMOGENEOUS POLYNOMIALS WITH A MULTIPLICATION THEOREM*

BY

O. C. HAZLETT

1. **Relation to the literature.** If $f(x)$ is any homogeneous polynomial in the n variables, $x_1, \cdots, x_n$, such that

$$f(x)f(\xi) = f(X) \tag{1}$$

where the X's are bilinear functions of the x's and ξ's,

$$X_k = \sum_{ij} \gamma_{ijk} x_i \xi_j \qquad (k = 1, \cdots, n), \tag{2}$$

then $f(x)$ is said to admit the composition (2).

One of the simplest such functions is $f(x) = x_1^2 + x_2^2$ for which

$$X_1 = x_1\xi_1 \pm x_2\xi_2, \quad X_2 = x_2\xi_1 \mp x_1\xi_2. \tag{3}$$

This was known to Diophantus.† Euler and Degen noticed that, similarly, a sum of n squares admits composition when $n=4$ or 8, and Hurwitz‡ proved that such is true only when $n=1, 2, 4, 8$. Two other well known examples are the multiplication theorem for determinants whose n^2 elements are independent variables and the norms of algebraic numbers. Moreover the composition of forms occurs in number theory.

Usually in number theory we are concerned with the composition of forms in a more general sense, viz,

$$f'(x)f''(\xi) = F(X) \tag{4}$$

where f^1, f^{11} and F are polynomials in the same number of variables and of the same degree, but not necessarily with the same coefficients. Dickson§ proved that the determination of three polynomials satisfying (4) can be reduced to the determination of one polynomial satisfying (1). Hence we can and shall restrict our attention to compositions (1).

---

* Presented to the Society, December 29, 1926; received by the editors in July, 1928.

† For the history and bibliography of the subject of composition of forms, see Dickson's *History of the Theory of Numbers*, vol. III, especially Chapters III, IV.

‡ For a simple exposition and an account of the history of this problem, see Dickson, Annals of Mathematics, (2), vol. 20 (1919), pp. 155–171, 297.

§ Comptes Rendus, Paris, vol. 172 (1921), pp. 636–640.

In another illuminating article,* Dickson pointed out that the determination of all polynomials $f(x)$ which satisfy (1) where the X_k are defined in terms of the x and ξ by (2) is equivalent to the determination of all polynomials $f(x)$ which satisfy (1) where $x\xi = X$ in the linear algebra E of order n whose multiplication table is

$$e_i e_j \equiv \sum_k \gamma_{ijk} e_k \qquad (i,j = 1, \cdots, n). \tag{5}$$

The determinant of the ξ_j in (2) is

$$\Delta(x) = \left| \sum_i \gamma_{ijk} x_i \right|,$$

the first (or right-hand) determinant of the number $x = \sum x_i e_i$ in the algebra E. When $f(x) = x_1^2 + x_2^2$ and the composition is given by (3), then the algebra (5) is $e_1e_1 = e_1$, $e_1e_2 = \mp e_2$, $e_2e_1 = e_2$, $e_2e_2 = \pm e_1$. When $f(x) = x_1^2 + x_2^2 + x_3^2 + x_4^2$ and the composition is $X_1 = \xi_1x_1 - \xi_2x_2 - \xi_3x_3 - \xi_4x_4$, $X_2 = \xi_2x_1 + \xi_1x_2 + \xi_4x_3 - \xi_3x_4$, $X_3 = \xi_3x_1 - \xi_4x_2 + \xi_1x_3 + \xi_2x_4$, $X_4 = \xi_4x_1 + \xi_3x_2 - \xi_2x_3 + \xi_1x_4$, then the corresponding algebra is real quaternions with $e_1 \equiv 1$, $e_2 = i$, $e_3 = j$, $e_4 = k$; whereas $f(x) = \sum x_i^2$ $(i = 1, \cdots, 8)$ admits composition under Cayley's algebra of eight units with the multiplication table

$$\begin{aligned} &e_i^2 = -1, \quad e_ie_j = -e_je_i \qquad (i,j = 2, \cdots, 8\,;\ i \neq j)\,; \\ &e_2e_3 = e_4, \quad e_2e_5 = e_6, \quad e_2e_7 = e_8, \quad e_3e_6 = e_8, \quad e_3e_5 = -e_7, \\ &e_4e_5 = e_8, \quad e_4e_6 = e_7. \end{aligned} \tag{6}$$

If we take the lower sign in the multiplication table of the binary algebra given above, we see that the sum of two squares admits composition under an algebra which is both associative and commutative, but in order to get an algebra under which the sum of four squares admits composition, we have to discard the commutative law, and finally, to get an algebra under which the sum of eight squares admits composition, we have to discard the associative law, also. This makes it not at all surprising that there is no linear algebra under which the sum of 16 or more squares admits composition.

Dickson proved the following:

THEOREM A. *If $f(x)$ is a homogeneous polynomial of degree d which admits a composition and is not expressible in fewer than n variables, we may assume that the coefficient of x_1^d is unity after applying a transformation cogredient in the x's, ξ's and X's. Then by applying a transformation on the x's which leaves*

* *Homogeneous polynomials with a multiplication theorem*, Comptes Rendus du Congrès International des Mathématiciens, Strasbourg, 1921.

$f(x)$ unaltered and a transformation on the ξ's which leaves $f(\xi)$ unaltered, we obtain for $f(x)$ a new composition $x\xi = X$ in a linear algebra having a principal unit e_1 for which $e_1e_j = e_j = e_je_1$ $(j = 1, \cdots, n)$.

Such a composition is said to be *normalized.*

THEOREM B. *If $f(x)$ admits normalized composition, any factor $E(x)$ of $f(x)$ admits the composition if it has unity as the coefficient of the term involving only x_1.*

THEOREM C. *If a polynomial admits normalized composition, each of its irreducible factors (with the convention of Theorem* B*) admits the composition and divides both $\Delta(x)$ and $\Delta'(x)$.*

After finding all forms in 2, 3 and 4 variables which admit composition in the field of complex numbers, Dickson remarks that he has not yet succeeded in deciding whether or not an irreducible form of degree 5 in 5 variables admits composition.

2. **General case.** Let $f(x)$ be a homogeneous polynomial in n variables of degree d which admits the composition (2) and which is not expressible as a polynomial in less than n variables. If we give to the n x's any set of values such that $f(x) = c \neq 0$, then $f(c^{1/d}\xi) = f(X)$. Since $c \neq 0$, then $c^{1/d}\xi_i (i = 1, \cdots, n)$ are n linearly independent variables. Hence, since f can not be expressed as a function of less than n variables (by a non-singular linear transformation), the n X's must be linearly independent and therefore $\Delta(X) \neq 0$. But $\Delta(X) = \Delta(x)\Delta(\xi)$ and therefore $\Delta(x) \neq 0$.

Next, give to the x's such a constant set of values that $f(x) = 0$ and let the ξ's be independent variables. Then $\Delta(x) = 0$. For if $\Delta(x) \neq 0$, we should have $f(X) \equiv f(x)f(\xi) \equiv 0$ when the X's are n linearly independent functions of the ξ's and therefore are n linearly independent variables. Since this is a contradiction (except in the trivial case when $f(x)$ is originally assumed to be identically zero), our assumption that $\Delta(x) \neq 0$ is false.

Combining the two results, we see that $\Delta(x) = 0$ is equivalent to $f(x) \equiv 0$. Similarly, $\Delta'(x) = 0$ is equivalent to $f(x) = 0$ and we have proved

THEOREM 1. *If $f(x)$ is a polynomial in n variables which admits composition and which can not be expressed as a polynomial in less than n variables, then $f(x)$ has precisely the same irreducible factors as $\Delta(x)$ and $\Delta'(x)$, the first and second determinants of the linear algebra defined by the composition.*

When $n < 5$, this was proved by Dickson.* Moreover, it is readily verified that this is true for the case when $f(x) = \sum x_i^2$ $(i = 1, \cdots, 8)$ and E is Cayley's

* His Strasbourg paper mentioned above, Theorem 5.

algebra of order eight whose multiplication table is given by (6).

COROLLARY. *The degree of an irreducible factor $f(x)$ of a function admitting normalized composition is a factor of the minimum number of variables necessary to express $f(x)$.*

For let E be the algebra defined by the composition and let the variables necessary to express f be denoted by $x_1, \cdots, x_n$. Then f admits composition under E and under the algebra obtained from E by deleting all numbers except those of the form $x_1e_1+\cdots+x_ne_n$, the equations defining this latter composition being (2). But the order of this algebra is n and thus Δ is of degree n; and, by Theorem 1, Δ is a power of $f(x)$. Thus the corollary is proved.

3. **Powers of an irreducible polynomial.** Combining Theorem 1 with Theorem B, we have

THEOREM 2. *Let $f(x)$ be a homogeneous polynomial in n variables admitting normalized composition (2) under the linear algebra E and such that no one of its factors can be expressed as a polynomial in less than n variables. Then $f(x)$ is a power of $\Delta(x)$, and $\Delta(x)$ is either irreducible or is a power of an irreducible function. Similarly for $\Delta'(x)$.*

We also readily prove Theorem 6 of Dickson's paper for general n (which he had proved only for $n<5$), namely

THEOREM 3. *If a form f in n variables admits normalized composition and has an irreducible factor I which is not expressible in less than n variables, then f is a power of I.*

For, since I admits the composition and is not expressible as a polynomial in less than n variables, then Theorem 1 applies and thus Δ is a power of I. But I is irreducible and f has no factors other than those of I. Hence f is a power of I.

THEOREM 4. *If $f(x)$ is a polynomial in n variables $(n>1)$ which admits composition and which can not be expressed as a polynomial in a smaller number of variables, then $f(x)$ is not irreducible of degree n in the field of complex numbers. On the other hand, if the field of definition is any field F, then the only such functions are the norms of numbers which are algebraic with respect to F.*

Without loss of generality, we shall prove this when the algebra is normalized. Since $f(x)$ is of degree n and is irreducible, then by Theorem 2, $\Delta(x)$ must be irreducible and identical with $f(x)$. Now the (first) characteristic equation of the general number x of the algebra is $\delta(x)\equiv\Delta(x-\omega\epsilon)=0$ where

ϵ is the principal unit or modulus and thus $\delta(x)$ is irreducible for the general number. Hence the entire algebra is generated by one number of the algebra and hence must be a field which is algebraic with respect to F of order n. But when F is the field of complex numbers, this is possible only when $n=1$.

THEOREM 5. *If $f(x)$ is a form in n variables $x_1, \cdots, x_n$ which admits composition and which vanishes only when $x_1=x_2=\cdots=x_n=0$, then the algebra defined by the composition is a division algebra; and conversely.*

Since $f(x)$ vanishes only when $x_1=x_2=\cdots=x_n=0$, it follows that no factor of $f(x)$ is expressible in less than n variables. For, if some factor were expressible in n_1 variables ($n_1<n$), we may without loss of generality assume these variables to be $x_1, \cdots, x_{n_1}$. Then $f(x)$ would vanish when $x_1=x_2=\cdots=x_{n_1}=0$ and $x_i (i>n_1)$ is arbitrary, which is a contradiction. Hence $f(x)$ is a power of $\Delta(x)$ and thus $\Delta(x)$ vanishes if and only if $x_1=x_2=\cdots=x_n=0$. Since, in any linear algebra, $\Delta(x)=0$ is a necessary and sufficient condition that x be a divisor of zero, it follows that the algebra defined by the composition has no divisors of zero and hence is a division algebra. This proves the first half of the theorem and the converse is proved by reversing the argument.

THEOREM 6. *If $f(x)$ is a power of an irreducible form in n variables which is not expressible in less than n variables, then the algebra defined by the composition is irreducible.*

For since $f(x)$ is a power of an irreducible polynomial in not less than n variables, so also is $\Delta(x)$ by Theorem 2. If E were reducible, then its basal numbers could be so chosen that they could be divided into two sets, the e's and $\bar{e}$'s, such that the product of two basal numbers from the same set is linearly dependent on the numbers of that set but such that the product of any two basal numbers from different sets is zero. Hence the first determinant, $\Delta(x)$, for the general number of the algebra is of the form $\Delta_1(x)\,\Delta_2(x)$, where $\Delta_1(x)$ is the first determinant for the general number of the subalgebra determined by the e's and $\Delta_2(x)$ is the first determinant for the general number of the subalgebra determined by the $\bar{e}$'s. Hence, if E were reducible, we should have a contradiction.

4. **Two or more irreducible factors.** Next, we shall consider the case when $f(x)$ is reducible but is not a power of an irreducible function. By Theorem 1, we may assume without loss of generality that $f(x)=\Delta(x)$. Let $f(x)=\Pi_i(f_i(x))^{p_i}$, where each f_i is an irreducible polynomial and where $f_1, f_2, \cdots$ are distinct. By Theorem 3, every one of the polynomials $f_1, f_2, \cdots$ can be expressed as a polynomial in less than n variables. Moreover, since

any factor of $f(x)$ (with the convention used in Theorem B) admits the composition, it follows from Theorem 1 that any factor of $f(x)$ can be expressed as a polynomial in less than n variables unless it contains each of the factors $f_i(x)$ with non-zero exponent.

As indicated above, let the composition be normalized with e_1 as the modulus of the algebra. Then, since $f_1(x)$ can be expressed as a polynomial in less than n variables, make a linear transformation on the variables x_i so that $f_1(x)$ is a polynomial in $x_1, \cdots, x_{n_1}$, in which n_1 is as small as possible. When we make the induced transformation on the basal numbers of the algebra, the modulus of the algebra is not in general e_1 and so the composition is not normalized, strictly speaking, although the composition still posseses the essential feature of normalized composition, namely, the algebra has a modulus or principal unit. Moreover, although the composition is no longer normalized, yet suitably chosen irreducible factors admit the composition, namely, the transforms of the original $f_i(x)$. Since the product of the rest of the $f_i(x)$ is expressible in less than n variables and since the totality of variables necessary in all the f_i taken together is n, then we may so choose the variables necessary to express $\phi_1 = \Pi f_i$ $(i>1)$ as $x_{n_1+1}, \cdots, x_n$ with, possibly, one or more chosen out of the set $x_1, \cdots, x_{n_1}$ and these latter we may, without loss of generality, assume to be $x_{i_1+1}, \cdots, x_{n_1}$ $(i_1>0)$.

First we prove that $f(x)$ is a product of polynomials in non-overlapping sets of variables. To prove this statement, we prove that the coefficient of any term in f_1 which contains the variables $x_k(k=i_1+1, \cdots, n_1)$ is necessarily zero, and proceed by induction.

In $f(x)$, set equal to zero every x_k where $k<i_1$. Then, if f_1 actually contain terms in the x_k alone where $k>i_1$, the result of making this substitution in f is a polynomial g in $n-i_1$ variables which is not identically zero and which admits the composition $g(x)g(\xi)=g(X)$ under the algebra B_1 obtained from E by deleting the numbers $x_1e_1+\cdots+x_{i_1}e_{i_1}$.* Since $g(x)$ and $\phi_1(x)$ admit composition under the same algebra B_1 and neither can be expressed in a smaller number of variables, then (by Theorem 1) $g(x)$ has the same irreducible factors as $\phi_1(x)$ and thus $\phi_1(x)$ has at least one factor which contains no variables from the set $x_j(j=n_1, \cdots, n)$. By regrouping the factors of $f(x)$, we can write $f(x)$ in the form $g_1(x)g_2(x)$ where $g_1(x)$ is a polynomial in the first n_1 variables and where $g_2(x)$ is a polynomial containing no factor in the first n_1 variables alone. If we now apply the preceding argument, we have a contradiction unless $g_2(x)$ were independent of the first n_1 variables. Hence

* Note that this is not the same as the statement that the totality of numbers of the form $\sum x_je_j(j\leqq i_1)$ form an algebra.

if $f_1(x)$ actually contain a term in the x_j $(j>i_1)$ alone, then f can be written in the form g_1g_2 where g_1 is a polynomial in the first n_1 variables and g_2 is a polynomial in the rest of the variables—two independent sets of variables. Moreover, since g_1 is a polynomial in n_1 variables that admits composition under the algebra obtained from E by deleting all except the first n_1 basal numbers, it follows (by Theorem 3 and the understanding about the variables) that g_1 is a power of f_1. This result is proved under the hypothesis that f_1 actually contains a term in the x_j alone $(j>i_1)$. Since this hypothesis leads to a contradiction, it follows that f_1 contains no term in these x_j alone. Here it is understood that f_1 can not be expressed in less than n_1 variables and that these x_j are the variables that actually appear in f_1 and also in the other f_k.

From this it follows that f_1 contains no term whatsoever in these $x_j(j>i_1)$, i.e., it can be expressed in less than n_1 variables. For if f_1 could not be expressed in less than n_1 variables, then we should be making an allowable transformation, that is, one that still expresses f_1 in the least number of variables, when we replace every x_k by $x_k+a_kx_{n_1}$ $(k<n_1)$. The coefficient of the term in x_{n_1} alone in the transformed f_1 would be $f_1(a_1, a_2, \cdots, a_{n_1-1}, 1)$. But, since f_1 is not identically zero, there is a set of numbers in the given field F such that this coefficient is not zero. Since this is a contradiction of the preceding paragraph, it follows that f_1 can be expressed in less than n_1 variables.

But this last conclusion is inconsistent with the significance of the notation for the variables, and thus the two sets of variables necessary to express f_1 and ϕ_1 are entirely independent of each other. Now ϕ_1 is a polynomial in $n-n_1$ variables which admits composition under E and hence under the algebra obtained from E by deleting the first n_1 basal numbers. Hence, by induction, it follows that f is a product of powers of irreducible polynomials in non-overlapping sets of variables.

This means that the composition (1) depends only on those γ_{ijk} where i, j, k are all chosen from the set $1, \cdots, m_1=n_1$ or all chosen from the second set $m_1+1, \cdots, m_2$, or all chosen from one of the sets $m_{p-1}+1, \cdots, m_p$ where $m_p=\sum n_i$ $(i=1, \cdots, p)$. The others may not be zero, conceivably; but they do not actually appear in $f(X)$ after like terms have been combined. For, since $f_1(x)f_1(\xi)=f_1(X)$, where only the first set of variables appears in f_1, it follows that although $X_k=\sum\gamma_{ijk}x_i\xi_j(k=1, \cdots, n_1)$, yet the terms in the variables ξ and x from sets other than the first must cancel in $f_1(X)$ and thus may be ignored so far as f_1 is concerned. Since these same variables do not appear in any other f_i, they do not affect f or the validity of the composition. Hence we may assume to be zero every γ_{ijk} in which all three subscripts

are not in the same set, and we have

THEOREM 7. *Let $f(x)$ be a homogeneous polynomial in n variables which admits composition and which can not be expressed in less than n variables. Then, if $f(x)$ is not the power of an irreducible polynomial, it is equivalent (under a linear transformation on the variables) to a product of powers of distinct irreducible forms, $f_i(x)$, where (1) f_i is a form in n_i variables which can not be expressed as a form in a smaller number of variables; (2) the various sets of variables are non-overlapping; (3) $\sum n_i = n$; (4) if E is any algebra defined by the composition, then $f(x)$ admits composition under the algebra obtained from E by replacing by zero every γ_{ijk} where i, j, k are not all from the same set (for details, see the proof of the theorem).*

5. **Applications.** In view of the last theorem, we see that, in order to determine all homogeneous polynomials in n variables which admit a composition and which are not expressible in less than n variables, we first determine all such irreducible polynomials of the type just described together with products of powers of homogeneous polynomials f_i in non-overlapping sets of variables such that each f_i admits composition in the variables appearing in f_i, in the manner described in detail in Theorem 7.

There are two possibilities: (a) $\Delta(x)$ is expressible as a polynomial in less than n variables; (b) $\Delta(x)$ is not so expressible. Since, in this paper, we are interested in finding all polynomials, $f(x)$, which admit composition, but are not concerned with the problem of determining all compositions (2) under which a given $f(x)$ admits composition, we can ignore the first possibility since a polynomial which can be expressed as a polynomial in a smaller number of variables can and will be considered as a polynomial in the smallest possible number of variables. For the same reason, if $f(x)$ is reducible, we shall consider merely the most convenient algebra defined by the composition, namely the reducible algebra determined in the proof of Theorem 7.

First, let $f(x)$ be a form of degree n in n variables with coefficients in the field F and let it be irreducible there. Then, by Theorem 4, $f(x)$ must be the norm of the general number of a field which is algebraic with respect to F of order n.

If $f(x)$ be of degree n but reducible in F, there are two possibilities: (1) $f(x)$ is a power of an irreducible polynomial in F; (2) $f(x)$ is the product of powers of two or more distinct forms, $f_1(x)$, $f_2(x)$, etc., each of which is irreducible in F but which is expressible in a smaller number of variables than n. In case (2), by Theorem 7, there is an algebra E defined by the composition on all n variables which is reducible and such that it is the direct sum of two or more algebras such that the assertion that $f(x)$ admits

composition under this algebra E is equivalent to the assertion that $f_i(x)$ admits composition under the algebra E_i for every i. Accordingly, to determine all such polynmials $f(x)$, we determine all such polynomials $f_1(x), f_2(x), \cdots, f_k(x)$ such that (i) $f_1(x)$ is a polynomial in n_1 variables, but not in a smaller number of variables, with similar conventions about $f_2(x)$, etc.; (ii) if the variables necessary for $f_1(x)$ be denoted by x with subscripts ranging from 1 to n_1 inclusive, and if the variables necessary for $f_2(x)$ be denoted by x with subscripts ranging from n_1+1 to n_1+n_2 inclusive with similar notations for the variables necessary for the remaining $f(x)$, then $n_1+n_2+\cdots+n_k=n$.

Third, let $f(x)$ be reducible, but the power of an irreducible polynomial, $\phi(x)$, of degree d. Then d is a factor of n, the number of variables needed. If the algebra defined by the composition is associative, then E must be semi-simple (that is, containing no invariant nilpotent sub-algebra). For if E is not semi-simple, then $E=S+N$ where S is a semi-simple algebra and N is a maximal invariant nilpotent sub-algebra, and it is well known that the characteristic determinant $\Delta(x)$ formed for the general number x of the algebra E is the product of $\Delta_s(x)$, the characteristic determinant of the component of x in S, by a determinant $D_n(x)$. Now it is readily proved (for any field F), that the irreducible factors of $D_n(x)$ are the same as the irreducible factors of $\Delta_s(x)$. Hence, when the algebra may be taken as associative, $\Delta(x)$ is expressible as a polynomial in less than n variables. If, however, the algebra may not be taken as associative, not sufficient is at present known about non-associative algebras to enable us to decide whether a similar result is true here, also.

Thus we see that if $f(x)$ is a form in n variables which admits composition, then we have the following results for small values of n:

THEOREM 8. *When $n=2$, $f(x)$ is a power of an irreducible form $\phi(x)$ of degree 2, where $\phi(x)$ is the norm of the general number of an algebraic field of order 2 with respect to F, or $f(x)$ has to be a product of powers of 2 linearly independent linear forms. When F is C, the field of complex numbers, $f(x)$ has to be a product of powers of 2 linearly independent linear forms.*

THEOREM 9. *When $n=3$, $f(x)$ has to be a power of an irreducible form $\phi(x)$ of order 3, where $\phi(x)$ is the norm of the general number of an algebraic field of order 3 with respect to F; or a product of powers of an irreducible form $\psi(x)$ of degree 2 in 2 variables, x_1, x_2, where $\psi(x)$ is the norm of the general number of a field which is algebraic of order 2 with respect to F, and a linear form in a third variable, x_3; or it has to be a product of powers of 3 linearly independent linear forms in 3 variables.*

THEOREM 10. *When $n=4$, $f(x)$ is one of the following:*

(i) *a power of an irreducible form $\phi(x)$ of degree 4, where $\phi(x)$ is the norm of the general number of a field which is algebraic of order 4 with respect to F;*

(ii) *a power of an irreducible form in 4 variables of degree 2;*

(iii) *a product of powers of an irreducible cubic in 3 variables, x_1, x_2, x_3, and the linear form, x_4;*

(iv) *a product of powers of two irreducible quadratics in 2 variables, the first in x_1, x_2 and the second in x_3, x_4;*

(v) *a product of powers of an irreducible quadratic in 2 variables, x_1, x_2, and of two linear forms, x_3, x_4;*

(vi) *a product of powers of the 4 linear forms, x_1, x_2, x_3, x_4. Each form mentioned above which is irreducible of degree k in k variables is the norm of the general number of a field which is algebraic with respect to F of order k.*

When F is the field of complex numbers, $f(x)$ has to be a power of an irreducible quadratic in 4 variables or a product of powers of 4 linearly independent linear forms.

THEOREM 11. *When $n=5$, $f(x)$ has to be one of the following:*

(i) *a power of an irreducible form $\phi(x)$ of degree 5, which is the norm of the general number of a field which is algebraic of order 5 with respect to F;*

(ii) *a product of powers of irreducible forms, $f_i(x)$, where each $f_i(x)$ is an irreducible form in less than 5 variables which admits composition, such that the variables actually necessary for any two or more $f_i(x)$ are linearly independent, and such that the total number of variables necessary is 5.*

In general, when n is a prime, $f(x)$ has to be a power of an irreducible form $\phi(x)$ of degree n, which is the norm of the general number of a field which is algebraic of order n with respect to F, or it has to be a product of powers of irreducible forms $f_i(x)$ in less than n variables which admits composition, such that the variables necessary for any two or more $f_i(x)$ are linearly independent and such that the total number of variables necessary is n.

When n is composite, $f(x)$ has to be of one of the two last mentioned types or it is a power of an irreducible form of degree q in n variables where q is a divisor of n.

UNIVERSITY OF ILLINOIS,
URBANA, ILL.

GENERALIZED FACTORIAL SERIES*

BY

TOMLINSON FORT

In this paper I consider series of the type

$$(1) \qquad c_0 + \sum_{n=1}^{\infty} c_n[\lambda_1\lambda_2 \cdots \lambda_n][(z+\lambda_1)(z+\lambda_2)\cdots(z+\lambda_n)]^{-1}$$

where $z=x+yi$ is a complex variable and the c_n's a sequence of complex constants. Unless otherwise specified $\lambda_n=\rho_n+\sigma_n i$, where $\rho_n\to\infty$, $\sum_{n=1}^{\infty}(\rho_n)^{-1}$ diverges and $\sigma_n/\rho_n\to 0$.

Probably the most interesting case is when λ_n is real. This was treated by Landau in what is now a classical memoir.† Theorem V of the present paper, even for the case that λ_n is real, is a wide extension of Landau's most fundamental result. He proves uniform convergence over a finite region, whereas it is proved here for a sectorial region‡ bounded by polynomial curves of arbitrary degree. For complex values of λ_n Theorem IV does not require that $\sum_{n=1}^{\infty}(\sigma_n/\rho_n)^2$ converge and in this respect is more general than V. Theorems VI and VII on absolute convergence seem also to be new. Regions of absolute convergence have been established§ but not so general as those found here. In §3 so-called equiconvergence theorems are proved for (1) and related exponential series.‖ It is believed that the results given there have not previously been obtained.

For an excellent introduction to the literature see the article in the Encyclopaedia¶ by Hilb and Szász.

1. An important rôle in several sections of this paper is played by the following lemma from the general theory of infinite series. Proof of the lemma is easy and is not given here.

---

* Presented to the Society, February 25, 1928; received by the editors April 27, 1928.

† Münchener Berichte, vol. 36 (1906), p. 151–218.

‡ Compare Pincherle, Rendiconti del Circolo Matematico di Palermo, vol. 37 (1914), pp. 379–390.

§ Pincherle, loc. cit., p. 386, also F. Nevanlinna, Annales Academiae Scientiarum Fennicae, (A), vol. 18, No. 3, p. 8.

‖ Compare Landau, loc. cit., p. 201; Pincherle, loc. cit., p. 386; Hille, Annals of Mathematics, (2), vol. 25, pp. 276–278.

¶ Encyklopädie der Mathematischen Wissenschaften, Band II_3 H_8, pp. 1268–72.

LEMMA 1. *Hypotheses*: (a) $a_n(z)$ *and* $b_n(z)$, $n=1, 2, \cdots$, *are defined at all points of a region,* R; (b) $\sum_{n=1}^{\infty} a_n(z)$ *converges uniformly over* R; (c) *there exist positive constants* n' *and* M *such that when* $m \geqq n'$, $\sum_{n=m}^{n} |\Delta b_n(z)| < M$; (d) *there exists a positive constant* M' *such that* $|b_{n'}(z)| < M'$.

Conclusion: $\sum_{n=1}^{\infty} a_n(z)\, b_n(z)$ *converges uniformly over* R.

The part of $b_n(z)$ will be played in the fundamental theorems of this section by

$$b_n^{(\lambda)}(z_0, z) = (z_0 + \lambda_1)(z_0 + \lambda_2) \cdots (z_0 + \lambda_n)[(z_0 + \lambda_1)(z + \lambda_2) \cdots (z + \lambda_n)]^{-1}.$$

We proceed to prove preliminary theorems (I, II, and III) relative to $b_n^{(\lambda)}(z_0, z)$.

PRELIMINARY THEOREM I. *If* $x_0 \geqq 0$ *and* $\rho_i > 0$ *for all values of* i *and in case* $\sum_{n=1}^{\infty} (y_0+\sigma_n)^2 (x_0+\rho_n)^{-2}$ *converges, then* (1) $b_n^{(\lambda)}(z_0, z)$ *remains uniformly finite in* z *over the half-plane, defined by* $x \geqq x_0$, (2) *given any positive constants* a *and* b *and positive integer* m *there exists a corresponding constant* N, *such that*

$$|b_m^{(\lambda)}(z_0, z)| < N/\{a(x - x_0)^m + b\}$$

over the half-plane, defined by $x \geqq x_0$.

We have

$$|b_n^{(\lambda)}(z_0, z)| = \frac{|z_0 + \lambda_1| \cdots |z_0 + \lambda_n|}{(x_0 + \rho_1) \cdots (x_0 + \rho_n)} \cdot \frac{(x + \rho_1) \cdots (x + \rho_n)}{|z + \lambda_1| \cdots |z + \lambda_n|} \cdot b_n^{(\rho)}(x_0, x)$$

$$\leqq \{[1 + ([y_0 + \sigma_1]/[x_0 + \rho_1])^2] \cdots [1 + ([y_0 + \sigma_n]/[x_0 + \rho_n])^2]\}^{1/2} \cdot b_n^{(\rho)}(x_0, x) \leqq \overline{N} b_n^{(\rho)}(x_0, x) \leqq \overline{N}.$$

Also

$$|b_m^{(\lambda)}(z_0, z)| \leqq \overline{N} b_m^{(\rho)}(x_0, x) \leqq \overline{N}(x_0 + \rho_1) \cdots (x_0 + \rho_m)/\{x^m + \rho_1 \cdots \rho_m\}$$

$$\leqq c\overline{N}(x_0 + \rho_1) \cdots (x_0 + \rho_m)/\{cx^m + c\rho_1 \cdots \rho_m\} \leqq N/\{a(x - x_0)^m + b\}$$

if

$$c > a, \quad c\rho_1 \cdots \rho_m > b \text{ and } N > c\overline{N}(x_0 + \rho_1) \cdots (x_0 + \rho_m).$$

PRELIMINARY THEOREM II. *If* $x_0 \geqq 0$ *and* $\rho_i > 0$ *for all values of* i, *and in case* $\sum_{n=1}^{\infty} (y_0+\sigma_n)^2 \cdot (x_0+\rho_n)^{-2}$ *converges; then* $\sum_{n=m}^{n} |\Delta b_n^{(\lambda)}(z_0, z)|$ *remains uniformly finite over the sectorial region defined by* $x \geqq x_0$ *and* $|y - y_0| \cdot (x - x_0)^{-1} \leqq a(x - x_0)^m + b$, *where* a *and* b *are any positive constants.*

We have

$$|\Delta b_n^{(\lambda)}(z_0,z)| = \frac{|z_0 - z|}{|z+\lambda_{n+1}|}\cdot|b_n^{(\lambda)}(z_0,z)|$$

$$\leqq -\overline{N}\frac{|z-z_0|}{x-x_0}\cdot\frac{x+\rho_{n+1}}{|z+\lambda_{n+1}|}\cdot(x_0-x)b_n^{(\rho)}(x_0,x)\cdot\frac{1}{x+\rho_{n+1}} \tag{2}$$

$$\leqq -\overline{N}\frac{|z-z_0|}{x-x_0}\Delta b_n^{(\rho)}(x_0,x).$$

From this

$$\sum_{n=m}^{n}|\Delta b_n^{(\lambda)}(z_0,z)| \leqq \overline{N}\frac{|z-z_0|}{x-x_0}(b_m^{(\rho)}(x_0,x) - b_n^{(\rho)}(x_0,x))$$

$$\leqq \overline{N}\frac{|z-z_0|}{x-x_0}b_m^{(\rho)}(x_0,x) \leqq \frac{|z-z_0|}{x-x_0}\cdot\frac{N}{a_1(x-x_0)^m+b_1} \leqq N$$

if $|z-z_0|\cdot(x-x_0)^{-1}<a_1(x-x_0)^m+b_1$. Choose a_1, and b_1, so that $a_1(x-x_0)^m+b_1>\{[a(x-x_0)^m+b]^2+1\}^{1/2}$, and the theorem follows.

PRELIMINARY THEOREM III. *If $x_0>0$ and $\rho_i>0$ for all values of i, $\sum_{n=1}^{n}|\Delta b_n^{(\lambda)}(z_0,z)|$ remain uniformly finite over the sector, S, defined by $x\geqq x_0$, and $|y-y_0|\cdot(x-x_0)^{-1}\leqq\tan\omega$, where ω is a fixed positive angle less than $\pi/2$.*

We have

$$|\Delta b_n^{(\lambda)}(z_0,z)| = \frac{|z-z_0|}{|z_1+\lambda_{n+1}|-|z_0+\lambda_{n+1}|}\cdot\Delta|b_n^{(\lambda)}(z_0,z)|.$$

We proceed to show that it is possible to choose a fixed N and M so that, when $n>M$,

$$\left|\frac{|z-z_0|}{|z+\lambda_{n+1}|-|z_0+\lambda_{n+1}|}\right| < N \tag{3}$$

over S. Replace z by $x+yi$, z_0 by x_0+y_0i, and λ_{n+1} by $\rho_{n+1}+\sigma_{n+1}i$. Then by elementary algebra we find the following inequality equivalent to (3):

$$\begin{aligned}
&4\rho_{n+1}^2\left[N^4-N^2\left(1+\left(\frac{y-y_0}{x-x_0}\right)^2\right)\right]+4\rho_{n+1}\left[N^4-N^2\left(1+\left(\frac{y-y_0}{x-x_0}\right)^2\right)\right](x+x_0)\\
&\quad+N^4(x+x_0)^2+N^4\left[\frac{y-y_0}{x-x_0}(y+y_0+2\sigma_{n+1})\right]^2\\
&\quad+4N^4\rho_{n+1}\left[\frac{y-y_0}{x-x_0}(y+y_0+2\sigma_{n+1})\right]\\
&\quad+2N^4\left[\frac{y-y_0}{x-x_0}(y+y_0+2\sigma_{n+1})\right](x+x_0)\\
&\quad-2N^2[x^2+y^2+x_0^2+y_0^2+2\sigma_{n+1}(y+y_0)+2\sigma_{n+1}^2]\left[1+\left(\frac{y-y_0}{x-x_0}\right)^2\right]\\
&\quad+[(x-x_0)^2+(y-y_0)^2]\left[1+\left(\frac{y-y_0}{x-x_0}\right)^2\right]>0,
\end{aligned} \tag{4}$$

which holds if

$$4\rho_{n+1}^2[N^4 - N^2\sec\omega] + 4\rho_{n+1}[N^4 - N^2\sec\omega](x + x_0)$$
$$+ N^4(x + x_0)^2 - 8N^4\rho_{n+1}(\tan\omega)\cdot[\,|\,y_0\,| + |\,\sigma_{n+1}\,|\,]$$
$$- 4N^4(|\,y_0\,| + |\,\sigma_{n+1}\,|\,)(\tan\omega)(x + x_0)$$
$$- 2N^2[x^2 + y^2 + x_0^2 + y_0^2 + 2\sigma_{n+1}(y + y_0) + 2\sigma_{n+1}^2]\sec\omega > 0.$$

Replace y by $|y_0| + m(x - x_0)$ and rearrange the terms. We get

$$\{4\rho_{n+1}^2[N^4 - N^2\sec\omega] - 8N^4\rho_{n+1}(\tan\omega)(|\,y_0\,| + |\,\sigma_{n+1}\,|\,)$$
$$- 4N^2[y_0^2 - mx_0|\,y_0\,|\,]\sec\omega - 4N^2\sigma_{n+1}^2\sec\omega\}$$
$$+ \{N^4(x + x_0)^2 - 2N^2[x^2(1 + m^2) + x_0^2(1 + m^2) + 2mx|\,y_0\,|\,]\sec\omega\}$$
$$+ \{4\rho_{n+1}(N^4 - N^2\sec\omega)(x + x_0) + 4N^2m^2x_0x\sec\omega$$
$$- 4N^2|\,\sigma_{n+1}\,|\,(m(x - x_0) + 2|\,y_0\,|\,)\sec\omega$$
$$- 4N^4(|\,y_0\,| + |\,\sigma_{n+1}\,|\,)(\tan\omega)(x + x_0)\} > 0.$$

Now choose $N > 1$ so large that the second brace is positive for all values of $x > x_0$. Then choose M so large that when $n \geqq M$, the first and third braces are positive for all values of $x > x_0$. Then, when $n \geqq M$,

$$\sum_{n=1}^{n} |\,\Delta b_n^{(\lambda)}(z_0, z)\,| < N\left|\sum_{n=1}^{n}\Delta\,|\,b_n^{(\lambda)}(z_0, z)\,|\right|$$
$$= N\,|\,|\,b_{n+1}^{(\lambda)}(z_0, z)\,| - |\,b_1^{(\lambda)}(z_0, z)\,|\,|\,.$$

We must now show that $|b_n^{(\lambda)}(z_0, z)|$ remains uniformly finite in z over S as n varies. Consider the last factor

$$\left|\frac{z_0 + \lambda_m}{z + \lambda_n}\right| = \left[\frac{(x_0 + \rho_n)^2 + (y_0 + \sigma_n)^2}{(x + \rho_n)^2 + (y + \sigma_n)^2}\right]^{1/2}.$$

This is less than 1 in case

$$-(x_0 + x) - 2\rho_n < \frac{y - y_0}{x - x_0}(y + y_0) + 2\sigma_n\frac{y - y_0}{x - x_0}.$$

But

$$((y - y_0)/(x - x_0))(y + y_0) + 2\sigma_n((y - y_0)/(x - x_0))$$
$$\geqq -2|\,y_0\,|\tan\omega - 2|\,\sigma_n\,|\tan\omega,$$

and

$$-(x_0 + x) - 2\rho_n < -2|\,y_0\,|\tan\omega - 2|\,\sigma_n\,|\tan\omega$$

if

$$\rho_n > (|\,y_0\,| + |\,\sigma_n\,|)\tan\omega.$$

This is true for sufficiently large n, independently of z. This completes the proof of the theorem.

Definition of deleted region. We shall often have occasion to speak of a region from which neighborhoods of those points $-\lambda_1, -\lambda_2, \cdots$, which lie in it, have been removed. Such regions will be referred to as deleted regions.

THEOREM IV. *If* (1) *converges when* $z=z_0$, *it converges uniformly over the deleted sector, S, defined by* $x \geqq x_0$ *and* $|y-y_0| \cdot (x-x_0)^{-1} \leqq \tan \omega$, *where* ω *is any fixed positive angle less than* $\pi/2$.

Assume $x_0>0$. There is no loss of generality in this. If the contrary is the case, suppose $x_0+\rho_i \geqq \delta>0$ when $i \geqq k$. Drop from consideration the first k terms of the series and factor $\lambda_1\lambda_2 \cdots \lambda_k/[(z+\lambda_1)(z+\lambda_2) \cdots (z+\lambda_k)]$ from each remaining term. If the resulting series converges uniformly over S, so does the original series. Replace z_0 by $\bar{z}_0=z_0+\lambda_k$ and λ_{k+m} by $\bar{\lambda}_m=\lambda_{k+m}-\lambda_k$. Make necessary changes in the C_n's. We also assume $\rho_i>0$ when $i \geqq 1$. If this is not true proceed as before rewriting the series a second time.

The proof now follows by Lemma 1, using Preliminary Theorem III and the fact that $b_1(z_0, z)$ remains finite over S.

THEOREM V. *If* $\sum_{k=1}^{\infty} (\sigma_n/\rho_n)^2$ *converges and if* (1) *converges at* $z=z_0$, *then it converges uniformly over any deleted region, R, defined by* $x \geqq x_0$ *and* $|y-y_0| \cdot (x-x_0)^{-1} \leqq a(x-x_0)^m+b$, *where a and b are any positive constants and m any fixed positive integer.*

Again assume $x_0>0$ and $\rho_i>0$ for all values of i. As in the proof of Theorem IV this involves no loss of generality.

We know that (1) converges uniformly over any sector, S, as defined in Theorem IV and hence converges at $\bar{x}_0=x_0+\delta$, $\delta>0$. Therefore, by Lemma 1, using Preliminary Theorem II, and the fact $b_n(z_0, z)$ remains finite over R, we see that (1) converges uniformly over the sectorial area, $\overline{R}$, defined by $x \geqq \bar{x}_0$ and $|y| \cdot (x-\bar{x}_0)^{-1} \leqq (x-x_0)^{m_2}$. Choose m_2 and ω so large that every point of R is in at least one of the regions S or $\overline{R}$. The theorem follows.

2. We prove the following theorem:

THEOREM VI. *If* (1) *converges absolutely at* $z=z_0$, *it converges uniformly absolutely over any deleted sectorial region, R, defined by* $x \geqq x_0$ *and* $|y-y_0| \cdot (x-x_0)^{-1} \leqq \tan \omega$.

We again assume $x_0>0$ and $\rho_i>0$, $i=1, 2, \cdots$.

Consider the ratio

$$b_n^{(\lambda)}(z_0, z) = |z_0+\lambda_1| \cdots |z_0+\lambda_n| / [|z+\lambda_1| \cdots |z+\lambda_n|].$$

We saw in the proof of Theorem III that this remains uniformly finite. Our theorem follows.

THEOREM VII. *If* (1) *converges absolutely at* $z=z_0$ *and if* $\sum_{n=1}^{\infty}(y_0+\sigma_n)^2\cdot(x_0+\rho_n)^{-2}$ *converges, then* (1) *converges uniformly absolutely over the deleted half-plane defined by* $x\geqq x_0$.

We proceed as in the previous proof except that we refer to Theorem I instead of Theorem III.

3. We now prove

THEOREM VIII. *If* s *is a positive integer and if* $\sum_{j=1}^{\infty}|\lambda_j|^{-s-1}$ *converges,* then series* (1) *converges at the same points as does*

$$\sum_{n=1}^{\infty}c_n e^{g_{1n}}, \quad \text{where} \quad g_{1n}=\sum_{\mu=1}^{s}(-1)^{\mu}\frac{z^{\mu}}{\mu}\sum_{j=1}^{n}\lambda_j^{-\mu}. \tag{5}$$

Series (1) *also converges absolutely in the same deleted region as* (5).

Consider z fixed and different from $-\lambda_1, -\lambda_2, \cdots$. Assume $|\lambda_j|\geqq\lambda>|z|$ for all values of j. By reasoning similar to that at the beginning of the proof of Theorem IV there is shown to be no loss of generality in this. We have

$$\frac{\lambda_j}{z+\lambda_j}=e^{-\log(1+z/\lambda_j)}=e^{\gamma},$$

$$\gamma=-\frac{z}{\lambda_j}+\frac{z^2}{2\lambda_j^2}-\cdots+(-1)^s\cdot\frac{z^s}{s\lambda_j^s}+(-1)^{s+1}\frac{z^{s+1}}{\lambda_j^{s+1}}\theta(z/\lambda_j),$$

where $|\theta(z/\lambda_j)|<\Theta$, independent of j. Hence

$$\frac{\lambda_1\cdots\lambda_n}{(z+\lambda_1)\cdots(z+\lambda_n)}=e^{g_{1n}+g_{2n}}, \tag{6}$$

where

$$g_{2n}=(-1)^{s+1}\cdot z^{s+1}\sum_{j=1}^{n}\theta(z/\lambda_j)\lambda_j^{-s-1}.$$

We now refer to Lemma 1. Let

$$a_n=c_n e^{g_{1n}}, \quad b_n=\{\lambda_1\cdots\lambda_n/[(z+\lambda_1)\cdots(z+\lambda_n)]\}e^{-g_{1n}}.$$

From (6)

$$b_n=e^{g_{2n}}.$$

---

* The proof of this theorem holds in case restrictions previously put on λ_j, except $|\lambda_j|\to\infty$, be removed; s also may be zero.

This remains finite. Moreover

$$|\Delta b_n| = |e^{g_{2n}}(e^{g_{3n}} - 1)|,$$

where

$$g_{3n} = (-1)^s \cdot z^{s+1}\lambda_{n+1}^{-s-1}\theta(z/\lambda_{n+1}).$$

Hence,

$$|\Delta b_n| = |b_n|\,|e^{g_{3n}} - 1| \leqq |b_n|(e^{g_{4n}} - 1) \leqq M|\lambda_{n+1}|^{-s-1},$$

where $g_{4n} = |z|^{s+1}\lambda_{n+1}^{-s-1}\Theta$ and M is independent of n. This, however, is the general term of a convergent series. All hypotheses of Lemma 1 are fulfilled, and we have proved that (1) converges at all points at which (5) converges. The converse is readily proved in the same way.

To prove the portion of the theorem relative to absolute convergence, simply take the ratio of corresponding terms. We get what we have just been calling b_n. This remains finite in both cases for a fixed z.

THEOREM IX. *Series* (1) *has the same abscissa of convergence* and the same abscissa of absolute convergence as the Dirichlet series*

$$\sum_{n=1}^{\infty} c_n e^{g_{5n}}, \qquad g_{5n} = -z\sum_{j=1}^{n}\lambda_j^{-1}. \tag{7}$$

In this theorem we assume λ_j to be real.

Let $z = x$ be real and as in the previous proof assume $\lambda_j > \lambda > |x|$. Then

$$\lambda_1 \cdots \lambda_n/[(x+\lambda_1)\cdots(x+\lambda_n)] = e^{g_{6n}}, \tag{8}$$

where

$$g_{6n} = -x\sum_{j=1}^{n}\lambda_j^{-1} + (x^2/2)\sum_{j=1}^{n}\theta(x/\lambda_j)\cdot\lambda_j^{-2}$$

and

$$0 < \theta(x/\lambda_j) < 1.$$

Suppose (7) to converge when $x = \alpha$. Let $\delta > 0$. When $x = \alpha + \delta$ we write (1) by means of (8) in the form

$$c_0 + \sum_{n=1}^{\infty} c_n e^{g_{7n}} e^{g_{8n}},$$

$$g_{7n} = -\alpha\sum_{j=1}^{n}\lambda_j^{-1},$$

$$g_{8n} = -\delta\sum_{j=1}^{n}\lambda_j^{-1} + \tfrac{1}{2}(\alpha+\delta)^2\sum_{j=1}^{n}\theta\lambda_j^{-2}.$$

* Schnee, Berlin Dissertation, Göttingen, 1908, p. 74 ff.

When n is sufficiently large, $e^{g_{9n}}$ is a positive decreasing monotonic sequence. To show this, increase n by 1. This multiplies $e^{g_{9n}}$ by a number less than 1. Then, by a well known* theorem, series (1) converges when $x=\alpha+\delta$. In other words, since s although positive is as small as desired, the abscissa of convergence of (1) is not greater than that of (7).

Next, suppose (1) to converge when $x=\alpha$. When $x=(\alpha+\delta)$ we write the general term of (7)

$$c_n e^{g_{9n}} e^{g_{10,n}} = c_n \{(\lambda_1 \cdots \lambda_n)/[(\alpha+\lambda_1)\cdots(\alpha+\lambda_n)]\} e^{g_{10,n}},$$

where

$$g_{9n} = -\alpha \sum_{j=1}^{n}(1/\lambda_j) + (1/2)\alpha^2 \sum_{j=1}^{n}\theta(\alpha/\lambda_j)(1/\lambda_j^2),$$

$$g_{10,n} = -\delta \sum_{j=1}^{n}(1/\lambda_j) - (1/2)\alpha^2 \sum_{j=1}^{n}\theta(\alpha/\lambda_j)(1/\lambda_j^2).$$

For sufficiently great n, $e^{g_{10,n}}$ is a positive decreasing monotonic sequence. Increasing n by unity multiplies it by a factor which is less than 1. Hence, reasoning as before, the abscissa of convergence of (7) is not greater than that of (1). This completes the proof.

Only trivial modifications of this proof are necessary to establish the portion of the theorem relative to absolute convergence.

---

* See, for example, Bromwich, *An Introduction to Infinite Series*, 2d edition, p. 58. Also readily proved by means of Lemma 1.

LEHIGH UNIVERSITY,
BETHLEHEM, PA.

ON EXTENDING A CONTINUOUS (1-1) CORRESPONDENCE (SECOND PAPER)*

BY
HARRY MERRILL GEHMAN

I. INTRODUCTION

In a previous paper† we have given a condition which is necessary and sufficient in order that it be possible to extend a continuous (1-1) correspondence of two plane continuous curves to a correspondence of their planes. In the present paper, we shall give a similar condition for the case of two plane point sets of a more general type. Theorem 2 of this paper includes as special cases the two theorems of our *First paper*.

DEFINITION. By an *E-set*, we shall mean a closed and bounded plane point set, each component‡ of which is a continuous curve, not more than a finite number of them being of diameter greater than any given positive number.

II. SOME PROPERTIES OF AN E-SET AND OF THE DOMAINS COMPLEMENTARY TO AN E-SET

1. If M is an E-set, we can let $M=m_0+M_1+M_2+\cdots$, where m_0 is totally disconnected and consists of all those components of M which consist of a single point, and for each value of i, M_i is a component of M. Note that m_0 may consist of an uncountable infinity of components of M, but $M-m_0$ consists of at most a countable infinity of components of M.

2. Of the domains complementary to an E-set M lying in a plane S, one is unbounded and the remainder are bounded. Let D denote the unbounded domain of $S-M$, and for each component M_x of M, let D_x denote the unbounded domain of $S-M_x$. The set D is a subset of each of the sets D_x, and hence is a subset of their common part.

Let N be the point set consisting of the sum of the sets $S-D_x$ corresponding to all the components M_x of M. The set N is an E-set, each component

---

* Presented to the Society, October 29, 1927; received by the editors in December, 1928.

† H. M. Gehman, *On extending a continuous* (1-1) *correspondence of two plane continuous curves to a correspondence of their planes*, these Transactions, vol. 28 (1926), pp. 252–265. We shall refer to this paper as *First paper*.

‡ A *component* or *maximal connected subset* of a point set M is a connected subset of M which is not a proper subset of any other connected subset of M.

of which has just one complementary domain. The set N therefore has just one complementary domain,* which is the same as the common part of all the sets D_x. Now since M is a subset of N, the set $S-N$ is a subset of $S-M$, and since $S-N$ is unbounded, it must be a subset of D. As we have shown above that D is a subset of $S-N$ (the common part of the sets D_x), it follows that $D=S-N$.

The boundary of D is the sum of the boundaries of the domains D_x corresponding to those components M_x of M which do not lie in a bounded complementary domain of any other component of M.

3. Let d denote a bounded domain of $S-M$. If M_x is any component of M, the domain d is a subset of some domain of $S-M_x$. If d is a subset of a bounded domain of $S-M_x$, then the diameter of M_x is greater than or equal to the diameter of d. Hence it follows from the definition of an E-set, that in the case of a finite number of components of M, d is a subset of a bounded complementary domain; in the case of the remaining components of M, d is a subset of the unbounded complementary domain.

Let $M_1, M_2, \cdots, M_n$ denote the components of M such that for each component M_i, there exists a bounded domain d_i of $S-M_i$ which contains d. Each domain d_i has an outer boundary† consisting of a simple closed curve J_i. The curves $J_1, J_2, \cdots, J_n$ are mutually exclusive since they belong to different components of M, and since each of them encloses d, it follows that of any pair of them, one encloses the other. Hence one of them, say J_k, is interior to each of the other simple closed curves of the set.

It follows by §2, after an inversion of the plane about a circle lying in d, that the outer boundary of d is the simple closed curve J_k, and that the entire boundary of d consists of the boundary of d_k and the boundary of each of the unbounded domains D_x complementary to a component M_x of M which lies in d_k and does not lie in a bounded complementary domain of any other component of M which lies in d_k.

4. We have shown in §3 that corresponding uniquely to each bounded domain d of $S-M$ is a bounded domain d_k complementary to a component M_k of M, such that d_k contains d and has the same outer boundary. Given any positive number ϵ, not more than a finite number of components of M are of diameter greater than ϵ, and each of these has not more than a finite number of bounded complementary domains of diameter greater than ϵ.

---

* H. M. Gehman, *Concerning acyclic continuous curves*, these Transactions, vol. 29 (1927), pp. 553–568. See Lemma F, p. 558.

† R. L. Moore, *Concerning continuous curves in the plane*, Mathematische Zeitschrift, vol. 15 (1922), pp. 254–260.

Hence there are not more than a finite number of domains of $S-M$ which are of diameter greater the ϵ.

5. By the methods used in §2 and the Lemma referred to there, and by means of an inversion of the plane in the case of a bounded domain, it follows that if P is a point of M which is a boundary point of a domain of $S-M$, then P is *accessible* from that domain. That is, if Q is any point of the domain, there exists an arc QP every point of which except P is a point of the domain.

III. Interior-preserving correspondences

Definition.* Let M and M' be E-sets lying in planes S and S' respectively, and let T be a continuous (1-1) correspondence such that $T(M)=M'$. Then we shall say that T *preserves interiors* if given any two simple closed curves J of M and J' of M' which are such that $T(J)=J'$ and which enclose respectively the subset K of M and the subset K' of M', then $T(K)=K'$.

Theorem 1. *If M and M' are E-sets lying in planes S and S' respectively, and if T is a continuous* (1-1) *correspondence such that $T(M)=M'$ and such that T preserves interiors, then if B is the boundary of a domain D of $S-M$, the set $T(B)$ is the boundary of a domain D' of $S'-M'$.*

In the case of the unbounded domain of $S-M$, the proof given in *First paper*, page 262, second paragraph, holds without change.

In the case of a bounded domain of $S-M$, the proof given in *First paper*, page 262, third paragraph, holds with the following changes: N consists of a countable set of trees and a (possibly uncountable) collection of components of M. Using the notation of §1 of Part II, we can denote this collection by $m_0+M_1+M_2+\cdots$. If D_i is the unbounded domain of $S-M_i$, let K_i denote $S-D_i$; and if D_i' is the unbounded domain of $S'-M_i'$, let K_i' denote $S'-D_i'$. Each component of the set $H=(L_1+L_2+\cdots)+m_0+(K_1+K_2+\cdots)$ is either one of the sets L_i, a point of m_0, or one of the sets K_i; similarly, each component of $H'=(L_1'+L_2'+\cdots)+m_0'+(K_1'+K_2'+\cdots)$ is one of the sets L_i', a point of m_0', or one of the sets K_i'. Furthermore since T preserves interiors, if a component of H is a set L_a, a point P of m_0, or a set K_b, then one component of H' is the set L_a', the point $T(P)=P'$ of m_0', or the set K_b', and conversely. Hence, as in *First paper*, not every point of I' (the interior of J') is a point of H', and it follows that $I'-H'$ consists of a single component, which is the required domain of $S'-M'$.

---

* *First paper*, p. 260.

IV. Correspondences which preserve sides of arcs in the same sense

Definitions. Let M and M' be E-sets lying in planes S and S' respectively, and let T be a continuous (1-1) correspondence such that $T(M)=M'$. We shall say that *T preserves sides of arcs*, if given any simple closed curve J in S containing an arc AB of M, then there exists a simple closed curve J' in S' containing $T(AB)=A'B'$ and such that if J encloses the subset K of M, then J' encloses $T(K)=K'$; and also if given any simple closed curve J' in S' containing an arc $A'B'$ of M', then there exists a simple closed curve J in S containing $T^{-1}(A'B')=AB$ and such that if J' encloses the subset K' of M', then J encloses $T^{-1}(K')=K$.* We shall say that *T preserves sides of arcs in the same sense*, if T preserves sides of arcs, and if given any two simple closed curves J_1 and J_2 in S containing arcs A_1B_1 and A_2B_2 of M respectively, then the corresponding simple closed curves J_1' and J_2' in S' can be constructed in such a way that if X_1, Y_1, Z_1 are three points of A_1B_1 and X_2, Y_2, Z_2 are three points of A_2B_2, such that the sense† $X_1Y_1Z_1$ on J_1 is the same as the sense $X_2Y_2Z_2$ on J_2, then the sense $T(X_1)T(Y_1)T(Z_1)$ on J_1' is the same as the sense $T(X_2)T(Y_2)T(Z_2)$ on J_2'; and similarly for any two simple closed curves in S' containing arcs of M'.

Theorem 2. *If M and M' are E-sets lying in planes S and S' respectively, and if T is a continuous* (1-1) *correspondence such that $T(M)=M'$ and such that T preserves interiors and preserves sides of arcs in the same sense, then T can be extended to a correspondence between the planes S and S', i.e., there exists a continuous* (1-1) *correspondence U, such that $U(S)=S'$, and such that for each point P of M, $U(P)=T(P)$.*

1. If every component of M is an arc or a point, and no interior point of an arc XY of M is a limit point of $M-XY$, then an arc in the plane S may be passed through M, and an arc in S' may be passed through M'.‡ If M and M' are subsets of arcs, it has been proved§ that *any* continuous (1-1) correspondence of M and M' can be extended to a correspondence of the

---

* In *First paper*, p. 253, we said that *sides are preserved under T*, if T has the above property. It seemed better to say here that *sides of arcs are preserved*. See Part V of this paper.

† J. R. Kline, *A definition of sense on closed curves in non-metrical plane analysis situs*, Annals of Mathematics, vol. 19 (1918), pp. 185–200, and J. R. Kline, *Concerning sense on closed curves in non-metrical plane analysis situs*, Annals of Mathematics, vol. 21 (1919), pp. 113–119.

‡ R. L. Moore and J. R. Kline, *On the most general plane closed point set through which it is possible to pass a simple continuous arc*, Annals of Mathematics, vol. 20 (1919), pp. 218–223.

§ R. L. Moore, *Conditions under which one of two given closed linear point sets may be thrown into the other one by a continuous transformation of a plane into itself*, American Journal of Mathematics, vol. 48 (1926), pp. 67–72.

planes, which proves Theorem 2 for sets of this type. The theorem just cited is a special case of Theorem 2, because if the E-sets M and M' are subsets of arcs, any continuous (1-1) correspondence of M and M' preserves interiors (vacuously) and preserves sides of arcs in the same sense.

2. In all other cases, we shall define, for convenience, a positive and a negative sense for points on simple closed curves in the planes S and S'.

If M contains any simple closed curves, we shall select an arbitrary simple closed curve J_1 of M and three arbitrary points A, B, C of J_1 and shall say that the sense XYZ on the simple closed curve J in the plane S is *positive*, if and only if it is the same as the sense ABC on J_1; otherwise the sense XYZ on J is *negative*. In the same way, in the plane S' we shall say that the sense $X'Y'Z'$ on the simple closed curve J' is *positive*, if and only if it is the same as the sense $T(A)T(B)T(C)$ on $T(J_1)$.

If M contains no simple closed curve, but contains an arc AB such that an interior point P of AB is a limit point of $M-AB$, then we shall construct an arbitrary simple closed curve J_1 containing the arc AB of M, and enclosing a subset K of $M-AB$ which has P for a limit point. Then we shall say that the sense XYZ on the simple closed curve J in the plane S is *positive*, if and only if it is the same as the sense APB on J_1. Since T preserves sides of arcs, there exists a simple closed curve J_1' in S' containing the arc $T(AB)$ and enclosing $T(K)$, and we shall define the sense $X'Y'Z'$ on the simple closed curve J' in the plane S' as *positive*, if and only if it is the same as the sense $T(A)T(P)T(B)$ on J_1'. With the help of Lemma A of *First paper* and Kline's definition of sense, it is easily seen that the positive sense in S' as defined thus is independent of the particular selection of the simple closed curve J_1'.

If in every application of the condition that T preserves sides of arcs in the same sense, we let J_1 be the fixed curve J_1 defined above, it follows that if three points of an arc of M which is a subset of the simple closed curve J_2, are in a certain sense (either positive or negative) on J_2, then the three points corresponding under T to the given points are in the same sense on the simple closed curve J_2', regardless of the particular selection of J_2'.

3. By Theorem 1, there exists a (1-1) correspondence between the domains of $S-M$ and those of $S'-M'$. We shall now show how to define the correspondence U for points of a pair of corresponding domains and their boundaries, this being done in such a way that for each point P of M, $U(P)=T(P)$.

For simplicity, we shall define U for points of the unbounded domains of $S-M$ and $S'-M'$, which we shall denote by D and D' respectively. A

similar method suffices for a pair of corresponding bounded domains, after an inversion of the plane.

4. As in §2 of Part II, let $N=S-D$ and let $N'=S'-D'$. If each component of N is considered as an element and each point of $S-N$ is considered as an element, the set of all such elements is topologically equivalent to the set of all points in a plane Σ.* The set of points in Σ corresponding to the components of N is a closed and totally disconnected set Γ. Similarly, if the components of N' and the points of $S'-N'$ are considered as elements, the set of all such elements is equivalent to the set of all points of a plane Σ', and the points of Σ' corresponding to the components of N' form a closed and totally disconnected set Γ'.

Since T preserves interiors, there exists a (1-1) correspondence V between the components of N and the components of N' such that a point P of M lies in a component N_x of N if and only if $T(P)$ lies in the component $V(N_x)$ of N'. The correspondence V considered as a correspondence between the points of Γ and the points of Γ' is continuous and (1-1). We have shown above that a continuous (1-1) correspondence between two closed and totally disconnected sets such as Γ and Γ', can be extended to their planes; that is, there exists a continuous (1-1) correspondence W such that $W(\Sigma)=\Sigma'$, and such that for each point P of Γ, $W(P)=V(P)$.

The correspondence W, considered as a correspondence between S and S', is a continuous (1-1) correspondence between the points of $S-N$ and the points of $S'-N'$. Furthermore W is such that if J is any simple closed curve in $S-N$, then J encloses the component N_x of N if and only if the simple closed curve $W(J)$ lying in $S'-N'$ encloses the component $V(N_x)$ of N'. This is equivalent to the statement that J encloses the point P of M if and only if $W(J)$ encloses the point $T(P)$ of M'.

5. For any given positive number ϵ, we can enclose† each component N_x of N by a simple closed curve J_x which has no points in common with N and which is such that every point of J_x and every point which is interior to J_x is at a distance less than ϵ from a point of N_x. Since N has the Heine-Borel property, some finite subset of this collection of simple closed curves encloses all points of N. Hence‡ there exists a finite set of simple closed curves $J_1, J_2, \cdots, J_k$ enclosing all points of N, each curve being exterior to

* See Theorem 25, p. 427, of the paper by R. L. Moore, *Concerning upper semi-continuous collections of continua*, these Transactions, vol. 27 (1925), pp. 416–428. See also H. M. Gehman, *Concerning acyclic continuous curves*, loc. cit., pp. 558–559.

† R. L. Moore, *Concerning the separation of point sets by curves*, Proceedings of the National Academy of Sciences, vol. 11 (1925), pp. 469–476.

‡ R. L. Moore and J. R. Kline, loc. cit., Lemma 1, p. 219.

every other one of the set, and every point lying on or interior to any one of the curves being at a distance less than ϵ from some point of N.

Under the correspondence W, these simple closed curves correspond in S' to a set of simple closed curves $J_1', J_2', \cdots, J_k'$ which enclose all points of N' and are mutually exterior. If it happens that the distance from some point lying on or interior to one of these curves to any point of N' is greater that or equal to ϵ, we can construct in S' a finite set of simple closed curves $C_1', C_2', \cdots, C_k'$ which enclose all points of N', are mutually exterior, are such that every point lying on or interior to one of the curves is at a distance less than ϵ from some point of N', and are such that each of the curves is interior to one of the curves J_i'. On account of this last condition, the set of simple closed curves $C_1, C_2, \cdots, C_k$ in S, where $C_i = W^{-1}(C_i')$, has all the properties of the set $J_1, J_2, \cdots, J_k$.

6. We shall now define the correspondence U for points of S which are not interior to any of the curves $C_1, C_2, \cdots, C_k$ as any continuous (1-1) correspondence such that $U(S) = S'$, $U(C_i) = C_i'$ for $i = 1, 2, \cdots, h$, and if $X_iY_iZ_i$ are three points of one of the curves C_i, then the sense $X_iY_iZ_i$ on C_i is the same as the sense $U(X_i)U(Y_i)U(Z_i)$ on C_i, *sense* being positive or negative as defined in §2. Note that the correspondence W can always be defined so as to have this last property, in which case we can let $U(P) = W(P)$ for all points P of S which are not interior to any of the curves C_i.

We shall next define U for points of D and of the boundary of D which are interior to the simple closed curve C, where C denotes any one of the curves $C_1, C_2, \cdots, C_k$.

7. If each component of M interior to C is either an arc or a point, and if no interior point of an arc XY of M is a limit point of $M - XY$, then just as in §1, we can construct an arc interior to C which contains all points of M which are interior to C, and hence there exists a continuous (1-1) correspondence U_1 such that (1) $U_1(S) = S'$, (2) for each point P of C, $U_1(P) = U(P)$, and (3) for each point P of M interior to C, $U_1(P) = T(P)$. We shall then let $U(P) = U_1(P)$ for all points of S interior to C.

8. If the subset of M interior to C is not of the above type, let M_1 be a component of M which is interior to C and which either contains an arc XY such that some interior point of XY is a limit point of $M - XY$, or contains a simple closed curve. If M_1 contains no simple closed curves, let A and B be the end points of the maximal arc of M_1 that contains the arc XY. If M_1 contains a simple closed curve, the component of the boundary of D which is a subset of M_1 contains a simple closed curve J_1. In this case, let A and B

be two points of J_1 which are not feet of trees* with respect to J_1 of the component of the boundary of D which contains J_1.

In either case, let E and F be two points of C. We can construct two non-intersecting arcs AE and BF in $S-C-M+A+B+E+F$.† Let $E'=U(E)$, $F'=U(F)$, $A'=T(A)$, and $B'=T(B)$, and in $S'-C'-M'+A'+B'+E'+F'$ we can construct two non-intersecting arcs $A'E'$ and $B'F'$.

Let G and H be two points which separate E and F on C and let $G'=U(G)$ and $H'=U(H)$. There are two and only two domains of $S-(C+AE+BF+M)$ which are interior to C and have boundary points on C. The boundary of one of these domains R_1 consists of the arc EGF of C, the arcs AE and BF, and a subset of M_1 containing an arc AKB which with the other three arcs forms a simple closed curve which is the outer boundary of the domain R_1. The other domain R_2 is bounded by the arc EHF of C, the arcs AE and BF, and a subset of M_1 containing an arc ALB which with the other three arcs forms a simple closed curve which is the outer boundary of the domain R_2. We have chosen A and B in such a way that either AKB and ALB are identical, or AKB and ALB have no interior points in common. It is also true that the boundary of R_1 (and that of R_2) contains the same points of M_1 regardless of how the arcs AE and BF are constructed.

Similarly in S', there are two and only two domains of $S'-(C'+A'E'+B'F'+M')$ which are interior to C' and have boundary points on C'. If the domain whose boundary contains G' is denoted by R_1', then the outer boundary of R_1' consists of $A'E'+E'G'F'+B'F'$ and one of the arcs $A'K'B'$, $A'L'B'$; the outer boundary of the other domain R_2' consists of $A'E'+E'H'F'+B'F'$ and the other one of the arcs $A'K'B'$, $A'L'B'$. The arcs $A'K'B'$, $A'L'B'$ either are identical or have no interior points in common. In the latter case, since the arcs $A'K'B'$ and $A'L'B'$ form a simple closed curve, it follows that the outer boundary of R_1' contains $A'K'B'$ because the following senses are either all positive or all negative: AKB on $AKBLA$, EGF on C, EGF on $EGFBKAE$, $A'K'B'$ on $A'K'B'L'A'$, and $E'G'F'$ on C'. It is also true that the subset of M_1' that lies in the boundary of R_1' (and that of R_2') is the same, regardless of how the arcs $A'E'$ and $B'F'$ are constructed. Since T preserves interiors and preserves sides of arcs in the same sense, it follows that a point P of M_1 is a point of the boundary of R_i $(i=1, 2)$, if and only if $T(P)$ is a point of the boundary of R_i'.

However it may happen that for a particular choice of the arcs AE, BF, $A'E'$, $B'F'$, a component M_2 of M lies in R_1, while $T(M_2)=M_2'$ lies in

* *First paper*, p. 262.

† Part II, §5.

R_2'. We shall now show how, keeping $A'E'$ and $B'F'$ fixed, a new pair of arcs AE and BF may be constructed so that a component M_2 of M lies in R_1 if and only if M_2' lies in R_1'.

9. Let us suppose that the arcs AE, BF, $A'E'$, $B'F'$ have been constructed. Let N_i' $(i=1, 2)$ denote the set consisting of all components of M' that lie in R_i', and let $N_i=T^{-1}(N_i')$. Each of the sets M_1+N_1, M_1+N_2, $M_1'+N_1'$, $M_1'+N_2'$ is closed.

We shall now show that if we denote by N_{1i} $(i=1, 2)$ the subset of N_1 that lies in R_i, then $N_{12}+A+B$ is closed. Since the set $M_1+N_1=M_1+N_{11}+N_{12}$ is closed, and no point of N_{11} is a limit point of N_{12}, it follows that M_1+N_{12} is closed. Since N_{12} is a subset of R_2 and $T(N_{12})$ is a subset of R_1', it follows that any point P of M_1 which is a limit point of N_{12} must be a boundary point of R_2, such that $T(P)$ is a boundary point of R_1'. If AKB and ALB have no interior points in common, A and B are the only points of M_1 which have this property. If $AKB=ALB$, every point of the arc has this property, but if an interior point of the arc AB of M_1 were a limit point of N_{12}, the correspondence T would not preserve sides of arcs in the same sense. Hence in both cases $N_{12}+A+B$ is closed.

Let us join H to L by an arc in R_2+H+L. This arc divides R_2 into two domains R_3 and R_4 whose boundaries contain the points A and B respectively. If we let N_{1i} $(i=3, 4)$ denote the subset of N_{12} which lies in R_i, then the sets $N_{13}+A$ and $N_{14}+B$ are closed.

Since A is a component of the closed set $N_{13}+A$, we can (by using Zoretti's theorem) let $N_{13}=K_1+K_2+\cdots$, where (1) K_i and K_j have no points in common, if $i\neq j$, (2) K_i and $N_{13}-K_i$ are closed, (3) $K_1+\cdots+K_i$ contains all components of N_{13} which contain points at a distance from A greater than $1/2^i$. Note that since T preserves interiors, N_{13} contains all components of M which lie in a bounded complementary domain of a component of N_{13}. Hence we can assume that the sets $K_1, K_2, \cdots$ also satisfy condition (4): if a component M_x of M lies in a bounded complementary domain of a component of K_i, then M_x is a component of K_i.

For each value of i, the sets K_i and $M-K_i$ are closed, and we can therefore, as in §5, enclose K_i by a finite number of mutually exterior simple closed curves $J_{i1}, J_{i2}, \cdots, J_{in_i}$, lying in R_3, of diameter less than $1/2^{i-2}$, such that the exterior of each one contains $M-K_i+J_{11}+\cdots+J_{1n_1}+J_{21}+\cdots+J_{i-1,n_{i-1}}$. For simplicity, we shall denote each of the curves J_{hk} just defined, by J_i, where $i=n_1+n_2+\cdots+n_{h-1}+k$. The point set consisting of $A+J_1+J_2+\cdots$ is closed; in fact it is an E-set. Also the set consisting of M, the simple closed curve C, the arcs AE, BF, HL, all the curves J_i, and any finite number of arcs, forms an E-set.

Let us next select for each value of i, a point Q_i of the curve J_i, and a point P_i whose distance from A is less than $1/2^i$, which is an interior point of the arc EA, and which is (for $i>1$) an interior point of the arc $P_{i-1}A$. We can then construct an arc P_iQ_i in $P_i+Q_i+R_3-(P_1Q_1+\cdots+P_{i-1}Q_{i-1})-(J_1+J_2+\cdots)$ of diameter less than $1/2^{h-2}$, where h is determined by the equation $i=n_1+\cdots+n_{h-1}+k$ given above. The set consisting of A and all the arcs P_iQ_i also forms an E-set.

No point of a given arc P_iQ_i is a limit point of M or of any subset of $\theta_i=(J_1+\cdots+J_{i-1}+J_{i+1}+\cdots)+(P_1Q_1+\cdots+P_{i-1}Q_{i-1}+P_{i+1}Q_{i+1}+\cdots)$. We can therefore enclose P_iQ_i by a simple closed curve C_i in R_1+R_3 whose exterior contains $M+\theta_i+C_1+\cdots+C_{i-1}$ and which is of diameter less than $1/2^{h-2}$, the relation between i and h being given above. The curve C_i contains two arcs $P_{i1}Q_{i1}, P_{i2}Q_{i2}$ such that (1) P_{i1} and P_{i2} are interior points of the arc EA lying between P_{i-1} and P_i and between P_i and P_{i+1} respectively, (2) Q_{i1} and Q_{i2} are points of J_i such that the sense $Q_{i1}Q_iQ_{i2}$ on J_i is the same as the sense EGF on $EGFBKAE$, (3) the interior of the simple closed curve formed by these two arcs and the arcs $P_{i1}P_{i2}$ of AE and $Q_{i1}Q_iQ_{i2}$ is a subset of R_3-M and contains no points interior to J_i.

The set consisting of $AE-(P_{11}P_{12}+P_{21}P_{22}+\cdots)+P_{11}Q_{11}+P_{21}Q_{21}+\cdots+P_{12}Q_{12}+P_{22}Q_{22}+\cdots+(J_1-Q_{11}Q_1Q_{12})+(J_2-Q_{21}Q_2Q_{22})+\cdots$ is an arc* from A to E which we may denote by $(AE)_1$. By our method of construction, the domain R_1 determined by using $(AE)_1$ in place of AE, contains $N_{11}+N_{13}$, while the distribution of points of N_2 between the domains R_1 and R_2 has not been changed. In the same way a new arc from B to F, which we may denote by $(BF)_1$, can be constructed so that the domain R_1 determined by $(AE)_1$ and $(BF)_1$ contains $N_{11}+N_{13}+N_{14}=N_1$. If any points of N_2 lie in the domain R_1 determined by $(AE)_1$ and $(BF)_1$, we can in the same way determine new arcs $(AE)_2$ and $(BF)_2$, such that the domain R_1 determined by them contains N_1, and the domain R_2 contains N_2.

10. Then as in *First paper*, pp. 264–265 and pp. 256–258, we can obtain a finite set of points $G_1=E, G_2, \cdots, G_m=F, \cdots, G_n$ lying in the order $G_1G_2\cdots G_m\cdots G_nG_1$ on the simple closed curve C, and a set of points $H_1=A, H_2, \cdots, H_m=B, \cdots, H_n$ of M_1, and a set of arcs $G_1H_1, \cdots, G_nH_n$, such that (1) the arc G_iH_i has no interior points in common with $C+M$, (2) the simple closed curve C_i formed by the arcs G_iH_i, $G_{i+1}H_{i+1}$, the arc G_iG_{i+1} of C that contains none of the points $G_1, \cdots, G_{i-1}, G_{i+2}, \cdots, G_n$, and an arc H_iH_{i+1} of the boundary of D, encloses no tree of $M_1-H_iH_{i+1}$

* This may be established with the aid of Theorem 2 of H. M. Gehman, *Some conditions under which a continuum is a continuous curve*, Annals of Mathematics, vol. 27 (1926), pp. 381–384.

which is of diameter greater than $\epsilon/9$, or whose corresponding set under T is of diameter greater than $\epsilon/9$, and (3) a point P of M is interior to C_i if and only if $T(P)$ is interior to the simple closed curve C_i' formed by the arcs $U(G_iG_{i+1})=G_i'G_{i+1}'$ of C', the arc $T(H_iH_{i+1})=H_i'H_{i+1}'$ of M_1, and arcs $G_i'H_i'$, $G_{i+1}'H_{i+1}'$ lying in $S'-C'-M'+G_i'+H_i'+G_{i+1}'+H_{i+1}'$.

Then by using the method of §5 and the method of *First paper*, p. 265 and pp. 258–259, we can express the points of D and its boundary which are interior to C as the sum of a finite number of domains containing no points of M, a finite number of domains of diameter less than ϵ, and the boundaries of these domains, the same being true for points of D' and its boundary which are interior to C'. By the method of §9, a point P of M lies in a domain in S, if and only if $T(P)$ lies in the corresponding domain in S'. The correspondence U is then defined as in *First paper*, p. 265 and pp. 259–260, so as to satisfy the conclusion of Theorem 2.

V. Logical relations between various kinds of correspondences

Definitions. Let M and M' be E-sets lying in planes S and S' respectively, and let T be a continuous (1-1) correspondence such that $T(M)=M'$. We shall say that T *preserves sides*, if given any simple closed curve J in S containing a connected* subset L of M, then there exists a simple closed curve J' in S' containing $T(L)=L'$ and such that if J encloses the subset K of M, then J' encloses $T(K)=K'$; and similarly for any simple closed curve J' in S' containing a connected subset L' of M'. We shall say that T *preserves sides in the same sense*, if T preserves sides, and if given any two simple closed curves J_1 and J_2 in S containing connected subsets L_1 and L_2 of M respectively, then the corresponding simple closed curves J_1' and J_2' in S' can be constructed in such a way that if X_1, Y_1, Z_1 are three points of L_1 and X_2, Y_2, Z_2 are three points of L_2, such that the sense $X_1Y_1Z_1$ on J_1 is the same as the sense $X_2Y_2Z_2$ on J_2, then the sense $T(X_1)T(Y_1)T(Z_1)$ on J_1' is the same as the sense $T(X_2)T(Y_2)T(Z_2)$ on J_2'; and similarly for any two simple closed curves in S' containing connected subsets of M'.

1. The following seven theorems, which follow from the definitions, give the complete set of logical relations connecting the five kinds of correspondences which have been defined in this paper.

Theorem 3. *If T preserves sides in the same sense, then T preserves sides.*

Theorem 4. *If T preserves sides of arcs in the same sense, then T preserves sides of arcs.*

* Since M is closed, L is necessarily either an arc or the curve J.

THEOREM 5. *If T preserves sides, then T preserves sides of arcs.*

THEOREM 6. *If T preserves sides in the same sense, then T preserves sides of arcs in the same sense.*

THEOREM 7. *If T preserves sides, then T preserves interiors.*

THEOREM 8. *If T preserves interiors and preserves sides of arcs, then T preserves sides.*

THEOREM 9. *If T preserves interiors and preserves sides of arcs in the same sense, then T preserves sides in the same sense.*

2. On account of Theorems 3, 6, 7, and 9, we can make a purely formal change in the statement of Theorem 2 by replacing "T preserves interiors and preserves sides of arcs in the same sense" by "T preserves sides in the same sense." However the conclusion of Theorem 2 no longer remains true if we replace this phrase by either "T preserves sides" or "T preserves sides of arcs in the same sense."

3. If the E-set M is connected, i.e., if M is a continuous curve, then it follows from the theorems of *First paper*, that the conditions (1) T preserves sides in the same sense, (2) T preserves sides of arcs in the same sense, (3) T preserves sides, (4) T preserves sides of arcs, are logically equivalent. Each of them implies that T preserves interiors but not conversely.*

---

* *First paper*, Lemma B, p. 261. Using the definitions of the present paper, Lemma B is *If M is connected and T preserves sides of arcs, then T preserves interiors.*

YALE UNIVERSITY,
NEW HAVEN, CONN.

A DETERMINATION OF ALL NORMAL DIVISION ALGEBRAS IN SIXTEEN UNITS*

BY

A. ADRIAN ALBERT

INTRODUCTION

The chief outstanding problem in the theory of linear associative algebras over an infinite field F is the determination of all division algebras. This problem is equivalent to that of the determination of all normal division algebras, or algebras A in which the only elements of A commutative with every element of A are the quantities of its reference field F.

The order of a normal division algebra is the square of an integer. All normal division algebras in 1^2†, 2^2‡, and 3^2§ units have been determined. In this paper all normal division algebras in 4^2 units, the next case, are determined and shown to be the algebras of Cecioni.‖

1. RESULTS¶ PRESUPPOSED

THEOREM I. *Every normal division algebra of order n^2 over a non-modular field has rank n.*

THEOREM II. *Every normal division algebra A of order n^2 contains an element q whose minimum equation is of degree n.*

Evidently the minimum equation of any element q of A is irreducible, and conversely, if an element of A is the root of an irreducible equation in F with leading coefficient one, then this equation is its minimum equation. We shall say that an element q of the algebra A is of rank k if the division algebra $F(q)$ is of rank k and thus if the minimum equation of q is of degree k. An equation will be said to be an irreducible equation in F if its coefficients are in F and if it is irreducible with respect to the field F.

---

* Presented to the Society, April 6, 1928; received by the editors in September, 1928.

† The only division algebras in one unit are fields.

‡ See Dickson, *Algebren und ihre Zahlentheorie*, p. 46.

§ Wedderburn, these Transactions, vol. 22 (1921), p. 132.

‖ Rendiconti del Circolo Matematico di Palermo, vol. 47 (1923), pp. 209–254.

¶ See Dickson, *Algebras and their Arithmetics*, German and English editions, for these results and their proofs.

THEOREM III. *The order of any sub-algebra of a division algebra is a divisor of the order of the main algebra.*

THEOREM IV. *Every element q of a normal division algebra A of order n^2 which is the root of an irreducible equation of degree n in F has the property that every element of A which is commutative with q is a polynomial in q with coefficients in F.*

THEOREM V. *The only elements of a normal division algebra A which are commutative with every element of A are elements of the reference field F.*

Evidently every rational function with coefficients in F of any element q of A is expressible as a polynomial in q with coefficients in F and is commutative with q. It may also be easily shown that any transform yqy^{-1} of an element q of A satisfies the minimum equation of q.

Hence the algebra A of order 16 over the non-modular field F contains an element satisfiying the reduced irreducible equation

$$\phi(\omega) \equiv \omega^4 + \alpha\omega^2 + \beta\omega + \gamma, \quad \alpha, \beta, \gamma \text{ in } F. \tag{1}$$

LEMMA 1. *If $A = BC$ where B and C are polynomials in an indeterminate ω over a division algebra D, and if $\omega - t$ is a right divisor of A but not of C, where t is in D, so that $C = Q(\omega - t) + R$, where $R \neq 0$ is independent of ω, then $\omega - y$ is a right divisor of B, where $y = RtR^{-1}$.*

THEOREM VI. *If A is a normal division algebra of order sixteen and if* (1) *is the minimum equation of an element q of A, then there exist further elements j, k, l of A which are transforms of q by elements of A such that*

$$\phi(\omega) \equiv (\omega - l)(\omega - k)(\omega - j)(\omega - q), \tag{2}$$

and also $\phi(\omega)$ is the product of the same factors permuted cyclically.

As a corollary of the proof of the above theorem we have the following theorem:

THEOREM VII. *If $\phi(\omega) = LM$, where L and M both contain ω, there exists a transform of q, $y = vqv^{-1}$, such that $\omega - y$ is not a right divisor of M and hence, if $M = Q(\omega - y) + R$, then $\omega - RyR^{-1}$ is a right divisor of L.*

2. REDUCTION TO THE DETERMINATION OF ALGEBRAS ASSOCIATED WITH THE GROUP G_8

Let A be any normal division algebra in sixteen units. If A contains an element i satisfying an irreducible equation in F such that, in the algebraic

sense, three of its distinct roots are rational functions of i and are distinct from i, then the algebra A is said to be of type Γ and is known.*

We shall now prove a theorem of fundamental importance in our determination.

THEOREM VIII. *If A is a normal division algebra in sixteen units containing an element q which satisfies the irreducible equation* (1) *with the property that not all of its ordinary complex roots are rational functions with coefficients in F of one of them, then A contains an element i satisfying the irreducible equation*

$$\psi(\omega) = \omega^4 + \delta\omega^2 + \epsilon = 0, \qquad \delta,\ \epsilon \text{ in } F. \tag{3}$$

By Theorem VI we may write the equation (2), where j, k, and l are transforms of q by elements of A. If $jq = qj$ then j is a polynomial in q with coefficients in F, by Theorem IV. Let $j = yqy^{-1}$. Then $j = \theta(q) = yqy^{-1}$. Therefore $\theta^2(q) = y\cdot\theta(q)\cdot y^{-1} = y^2qy^{-2}$. But $j \neq i$ since (1) is irreducible. Thus $\theta^2(q) \neq \theta(q) = j$. If $\theta^2(q) \neq q$ then the roots of (1) are q, j, $\theta^2(q)$, $-q-j-\theta^2(q)$, and are all polynomials in q. But this is impossible since then the ordinary complex roots of (1) are all rational functions of one of them with coefficients in F. Hence $q = \theta^2(q) = y^2qy^{-2}$. Therefore $y^2q = qy^2$ and, by Theorem IV, y^2 is a polynomial in q. But if y^2 is of rank four then $(1, y^2, y^4, y^6) = F(y^2)$ is an algebra of order four contained in $F(y)$, of order at most four, and hence is $F(y)$. Hence y is a polynomial in y^2 and consequently in q and is commutative with q. Hence, in this case, $j = yqy^{-1} = q$, a contradiction. Hence y^2 is not of rank 4 but is of rank 1 or 2 and satisfies $\omega^2+\delta\omega+\epsilon=0$, δ, ϵ in F, and thus $y^4+\delta y^2+\epsilon=0$, and y, in A, satisfies an equation of the desired type and is taken as the desired element if its rank is four. If not, then, since its rank is evidently not one, y is the root of an irreducible quadratic. (y could not be of rank 3 since A contains no sub-algebras of order 3, a non-divisor of 16.) Hence $y^2 = -\xi y+\eta$. If $\xi \neq 0$ this equation expresses y as a polynomial in y^2 and hence as a polynomial in q, which is not possible since $yq \neq qy$. Hence $\xi = 0$ and $y^2 = \eta$ in F. Let $z = qy$. Then $z^2 = qyqy = qjy^2 = qj\eta$. If z is of rank 2 then it satisfies $z^2 = \lambda z+\mu$. If $\lambda \neq 0$, this expresses $z = qy$ as a linear function of z^2 and thereby as a polynomial in q since j is a polynomial in q by hypothesis. This same relation gives $y = q^{-1}z$ as a polynomial in q which contradicts the hypothesis. Hence $\lambda = 0$. But $z^2 = \eta qj$. Hence $j = \kappa/q$ is a root of (1), where $\kappa = \mu/\eta \neq 0$ is in F. Substituting this value for the indeterminate in (1) and multiplying by q^4 we obtain $\gamma q^4+\beta\kappa q^3+\kappa^2q^2+\kappa^4=0$. Substituting the value of q^4 from (1) we obtain a linear combination of $1, q, q^2, q^3$ equal to zero and thus have the property that the coefficients are all zero. The coeffi-

* See Dickson, *New division algebras*, these Transactions, vol. 28 (1926), pp. 207–234.

cient of q^3 is $\beta\kappa=0$ and since $\kappa\neq 0$, $\beta=0$ and (1) is the desired equation. The only remaining case is thus the case where z is of rank 4. But $RzR^{-1}=-z$ where $R=j-q$ and $yj=qy$, $yq=jy$. Hence $-z$ is a root of the quartic equation satisfied by z and thus the equation satisfied by z has only even powers and is (3).

We shall prove, as the next step in our proof, the following lemma:

LEMMA 2. *If x is of rank 4 and $F(x)$ contains an element r of rank 2, then $F(x)$ and hence A contains an element i satisfying an equation of type* (3).

Let the equation of the element r be $r^2+\gamma_1 r+\gamma_2=0$, γ_1, γ_2 in F. Let $s=r+\gamma_1$. Then $s^2=\epsilon$ in F. The elements 1, s, x, sx are linearly independent for, if not, then $a+bx=0$, for a and b not both zero, where a and b are in $F(s)$. Thus if $b=0$ the equation becomes $a=0$, a contradiction, and if $b\neq 0$, then it has an inverse in the quadratic field $F(s)$, whence $x=c$ where $c=\lambda+\mu s$, an element of $F(s)$. But every element of $F(s)$ evidently satisfies a quadratic with coefficients in F, since the element s^2 is in F. Hence x is of rank 2, contrary to the hypothesis that x is of rank 4. Hence we may take the elements 1, s, x, sx as the basal units of $F(x)$. Therefore x^2, which is in $F(x)$, is expressible as a linear combination of the four units with coefficients in F and we obtain

$$x^2 = 2ax + b - a^2 \quad \text{or} \quad (x+a)^2 = b, \quad a, b \text{ in } F(s). \tag{4}$$

If the element $y=x+a$ is of rank two then b is in F, since then $y^2+\lambda y+\mu=0$ and thus $b+\lambda(x+a)+\mu=0$, or $\lambda x+(\lambda a+b+\mu)=0$. But 1, x, s are linearly independent. Hence $\lambda=0$ and $b=-\mu$ in F. Take the expression $a=\alpha_1+\alpha_2 s$. We have $[(x+\alpha_1)+\alpha_2 s]^2=\beta$ in F. The element $x+\alpha_1$ is of rank 4, since if not, then $\psi(x+\alpha_1)\equiv\theta(x)=0$ and x is the root of a quadratic in F, a contradiction. Let $x+\alpha_1=u$, of rank 4. Then $u^2+2\alpha_2 su+\alpha_2^2\epsilon=\xi$. Let $r=-2\alpha_1 s$, and $\beta-\alpha_2^2\epsilon=\gamma$. Then $u^2=ru+\gamma$ where $r^2=\xi$ of F, and $u^3=ru^2+\gamma u=r(ru+\gamma)=\gamma u=\xi u+\gamma u+\gamma r$, so that we have $u^4=(\xi+\gamma)u^2+\gamma(u^2-\gamma)=(2\gamma+\xi)u^2-\gamma^2$, an equation of the desired form. (The Greek letters of (1) are not these above.)

There remains only the case where y is of rank 4. We then have $b=\beta_1+\beta_2 s$, and obtain $y^4=\beta_1^2+2\beta_1\beta_2 s+\beta_2^2=\beta_1^2+\beta_2^2\epsilon+2(y^2+\beta_1)$, an equation with coefficients in F, and which is in the desired form. Thus the lemma is proved.

We shall now consider the remaining case, that where $jq\neq qj$. Equating the coefficients in (1) and (2) we have

$$q+j+k+l=0, \quad lk+(lj+lq)+(kj+kq)+jq=\alpha, \tag{5}$$

$$lkjq=\gamma, \quad lkj+lkq+ljq+kjq=-\beta, \tag{6}$$

from which, if we let $s=jq$, $t=q+j$, we obtain

$$\gamma s^{-1} + s = t + \alpha, \qquad \gamma s^{-1} = ts + \beta. \tag{7}$$

Hence t^2 is expressible as a polynomial in the element s of A with coefficients in F by the fact that s^{-1} is so expressible.

Case A. If t is of rank 4, then, if t^2 is of rank 2, by the lemma A contains an element of the desired kind. It is thus sufficient to treat the case t^2 of rank 4, since if t^2 is of rank one, the only other possible case, then t is the root of the quadratic $t^2=a$, a in F, a contradiction. Then $1, t^2, t^4, t^6$ are linearly independent and may be taken as the basal units of $F(t)$ whence t is expressible as a polynomial in t^2 and hence in the element s. Hence $st=ts$. Hence, since t is of rank 4, s is in $F(t)$. Hence, from $(7)_2$ and $(6)_1$ where $r=s^{-1}$, $r+s=t^2+\alpha$, $rs=\gamma$, and since all of the elements are commutative, we have, by elementary algebra, that r and s are roots of the equation $w^2-(t^2+\alpha)w+\gamma=0$ or thus $s^2=(t^2+\alpha)s-\gamma$. Also we obtain from $(7)_1$ by multiplying it on the right by t

$$rt + st = (t^2 + \alpha)t. \tag{8}$$

Adding (8) and the equation $st-rt=-\beta$ which is obtained from $(7)_2$ by the property that $st=ts$, we obtain

$$2st = t^3 + \alpha t + \beta \equiv z. \tag{9}$$

Thus $z-\beta=t(t^2+\alpha)$. Squaring equation (9) and applying the relation $s^2=s(t^2+\alpha)-\gamma$, we have $z^2-2\beta z=4t^2\gamma$, and replacing $z-\beta$ by its value $t^3+\alpha t$ in the equation

$$(z - \beta)^2 = 4t^2\gamma + \beta^2, \tag{10}$$

we have $t^2(t^2+\alpha)^2=4t^2\gamma+\beta^2$, a cubic in t^2, contrary to the hypothesis that t^2 is of rank four. Hence we need only consider the case t of rank two.

Case B. Here $t^2+\lambda t+\mu=0$. If $\lambda\neq 0$ then this equation expresses t as a function of t^2 and hence of s. Hence t is commutative with s and $(7)_2$ becomes $r-s=-\beta t^{-1}$. Subtracting this from $(7)_1$ we obtain the relation $2s$ equals a polynomial in t, or $s=\xi+\eta t$, since t is of rank 2. Hence $jq=\xi+\eta(j+q)$, $j(q-\eta)=\xi+\eta q$. But $q-\eta\neq 0$. Hence it has an inverse and j is expressible as a rational function of q, a contradiction, and $\lambda=0$, $t^2=f$, in F. But $t^2=(j+q)^2=q^2+qj+jq+j^2$. We have $j^2-q^2=f-qj-jq-2q^2=f-(qj+q^2+jq+q^2)=f-(qt+tq)$. But $t(qt+tq)=tqt+ft=tqt+tf=(qt+tq)t$. Hence t is commutative with j^2-q^2. If this element is of rank 4, then, by Lemma 2, $F(z)$, where $z=j^2-q^2$, contains t of rank 2 and hence an element of the desired type. If this is not so, then, if $z+t$ is of rank 4, the theorem is satisfied, since t is commutative with $z+t$ and hence is in $F(z+t)$ and is of rank 2.

The conclusion then follows by Lemma 2. Hence the only remaining case is that where z and $z+t$ are both of rank 2, where $zt=tz$. Let $z^2=\lambda z+\mu$, λ, μ in F. Let $(z+t)^2+\xi(z+t)+\eta=0$, ξ, η in F. Then we have that $z^2+2tz+t^2+\xi(z+t)+\eta=0$, whence $z(\lambda+\xi+2t)=-(\mu-f+\eta+\xi t)$. But $\lambda+\xi+2t\neq 0$ since t is not in F. Hence the above equation expresses z as a rational function of t and since t is of rank 2 we have $z=\rho+\sigma t$, ρ, σ in F. Hence we have, by the definition of z and t, that $j^2-i^2=\rho+\sigma(i+j)$ or $j^2-\sigma j=i^2+\sigma i+\rho$. Hence $y=j^2-\sigma j$ is commutative with i and is a polynomial in i. If y is of rank 2, then, by Lemma 2, the theorem is satisfied. If not, then y, in $F(j)$, is of rank 4 and hence the elements $1, y, y^2, y^3$ may be taken as a basis of $F(j)$ and hence j is expressible as a polynomial in y with coefficients in F and thereby as a polynomial in i with coefficients in F, a contradiction, since we assumed at the beginning of this section that $ij-ji\neq 0$. This proves the theorem.

If the normal division algebra A contains an element satisfying an equation of type (3) such that all of its roots are rational functions of one of them, then A is of type Γ. If A does not, then it contains an element satisfying an equation of type (3) such that not all of its roots are rational functions of one of them, by the theorem just proved. The algebras of type Γ are of two known kinds. They are either algebras of type D associated with an element satisfying a cyclic quartic, or are algebras of type E satisfying a non-cyclic abelian quartic or quartic with group G_4. Also every normal division algebra in sixteen units containing an element satisfying a cyclic quartic is an algebra of type D, and every normal division algebra containing an element satisfying a quartic with group G_4 is an algebra of type E. The necessary and sufficient condition that a quartic have group G_4 is that it be irreducible and that all of its roots be rational functions of one of them such that, if i and $\theta(i)$ are roots then $\theta^2(i)=i$, so that the equation is non-cyclic. We shall prove

THEOREM IX. *Every normal division algebra in sixteen units over a field F contains an element p satisfying a quartic with coefficients in F and group G_4. Hence every normal division algebra in sixteen units is an algebra of type E.*

Let A be a normal division algebra in sixteen units over F. Then, by the argument just made, A is of type Γ or A contains an element i satisfying the quartic equation

$$\phi(\omega) \equiv \omega^4 + \alpha\omega^2 + \beta = 0, \quad \alpha, \beta \text{ in } F, \tag{11}$$

irreducible in F, and such that not all of its roots are rational functions of one of them. Since i is a root of (11) so is $-i$. Applying Lemma 1 with $C=-i$ and $t=-i$ to $A=\phi(\omega)$ we have $R=-2i$ and $RtR^{-1}=-i$. Hence

$\phi(\omega)=Q(\omega)(\omega+i)(\omega-i)$. By Theorem VII there exists an element v of A such that $\omega-x$ is not a right divisor of $C=\omega^2-i^2$, where $x=viv^{-1}$. Hence, where the remainder on division of C by $\omega-x$ may be easily verified to be $R=x^2-i^2$, $\omega-RxR^{-1}$ is a right divisor of Q. Since Q is a polynomial in ω^2, $\omega+RxR^{-1}$ is a right divisor of Q and hence, by Lemma 1, $Q=(\omega+k)(\omega-k)$ where $k=ziz^{-1}$ and $z=Rv=(x^2-i^2)v$. We hence have the decomposition of (11) in the form (2) where here $l=-k$ and $j=-i$. Comparing coefficients in this decomposition with (11) we obtain the relations

$$k^2+i^2=-\alpha, \qquad k^2i^2=\beta. \tag{12}$$

Since $k=ziz^{-1}$, $zi=kz$, $zi^2=k^2z$, $z^2i^2=zzi^2=zk^2z=z(-\alpha-i^2)z=(-\alpha+\alpha+i^2)z^2=i^2z^2$. Obviously k^2 is a polynomial in i^2 and hence is commutative with i. Similarly i^2 is commutative with k. Also if $g(i)$ is any polynomial in i then $zg=g(k)z$. Let $z^2=s$, where s is in F. Then $z^2i=iz^2$. Let $y=ik-ki$, so that $iy=i^2k-iki=ki^2-iki=-yi$. Hence $iyi^{-1}=-y$. Also $z(ik-ki)=kzk-zki=kzziz^{-1}-zziz^{-1}i=ksiz^{-1}-siz^{-1}i=kiz-izi=(ki-ik)z=-yz$, so that $zyz^{-1}=-y$. The element $p=i^2+y$ is in the algebra consisting of all polynomials in y and i^2, with coefficients in F. This algebra is a field since $yi^2=-iyi=i^2y$. Hence, since A is of rank four, it is a field of order four or of order two. Its order is not two since y is not in the quadratic sub-field $F(i^2)$ for otherwise y would be a polynomial in i^2 and hence commutative with i, a contradiction of $yi=-iy$; while $y\neq 0$ since if $y=0$ then $ik=ki$ and k is a polynomial in i and the roots of (11) are all polynomials in i, contrary to hypothesis. Hence the field $F(i^2, y)$ is of order 4. The element p is not of rank 1 since it is obviously not in F. It is not of rank 2, for then $y^2=bp+c$, where b and c are in F, whence $i^4+2i^2y+y^2=bi^2+by+c$. But since $y^2i=yyi=-yiy=iy^2$ then y^2 is a polynomial in i. But $yg(i)=g(-i)y$, for any polynomial $g(i)$, and $yy^2=y^2y$. Hence y^2 is a polynomial in i^2. Hence, since y is not in $F(i^2)$, $2i^2=b$, and a contradiction is secured since i is the root of an irreducible quartic. Hence p is of rank 4. But $F(p)$ is then of order 4 and is contained in $F(i^2, y)$ of order 4. Hence $F(p)=F(i^2, y)$ and every polynomial in i^2 and y is a polynomial in p. Let $p_2=ipi^{-1}=i(i^2+y)i^{-1}=i^2-y$, $p_3=zpz^{-1}=zyz^{-1}+zi^2z^{-1}=-y+k^2$, $p_4=zp_2z^{-1}=ip_3i^{-1}=y+k^2$. These elements are transforms of the element p and hence are roots of the irreducible quartic satisfied by p. They are in the field $F(i^2, y)$ and hence are polynomials in p. Since z^2 and i^2 are both commutative with both i^2 and y they are commutative with p and hence $p_2^2(p)=p_3^2(p)=p_4^2(p)=p$. The elements p, p_2, p_3, p_4 are all distinct since $k^2\neq i^2$ and $y\neq -y$ while y is not a polynomial in i^2. Hence the quartic satisfied by p has the four distinct roots p, p_3, p_2, p_4 all of which are

polynomials in p, such that the iterative of the polynomial gives p and hence such that the group of the equation is G_4.

We shall next consider $z^2=s$ not in F. Let $z^2iz^{-2}=m$. If $m=i$ then z^2 is a polynomial $a(i)$. Since $zz^2=z^2z$, $a(i)=a(k)$. Let $a(i)=a_1+a_2i$, where a_1 and a_2 are in $F(i^2)$. Then $a_1+a_2i=a_1'+a_2'k$ where $a_1'=a_1(k^2)$ and $a_2'=a_2(k^2)$; both a_1' and a_2' are hence in $F(i^2)$. But k is not a polynomial in i. Hence $a_2'=0$. Hence $a_2=0$. Hence $a_1=a_1'$ which is true only if a_1 is in F, since $i^2\neq k^2$ and hence the coefficient of i^2 in a_1 must vanish. Hence s is in F, a contradiction. Hence $m\neq i$. Hence $j=m-i\neq 0$. If s is of rank four then the field $F(s)=F(z)$ since s and hence $F(s)$ is contained in $F(z)$, a field of order less than or equal to four. In this case z is a polynomial in s and hence is commutative with i^2 since s is. But z is not commutative with i^2. Hence s is not of rank four. But, by hypothesis, s is not in F and hence is not of rank 1. Hence the rank of s is 2 and s is a root of $w^2+2bw+c$ where b and c are in F. Let $t=s+b$. Hence $t^2=e$ in F. Since $t+b=s+2b=-cs^{-1}$, we have $jt+tj=(m-i)(s+b)+(s+b)(m-i)=ms-si+sm+2b(m-i)=ms-ms+(s+2b)m-i(s+2b)=-cs^{-1}(sis^{-1})+c(is^{-1})=-cis^{-1}+cis^{-1}=0$. Therefore $jt=-tj$. Hence $jtj^{-1}=-t$. Since $z^2i^2=i^2z^2$ and $ii^2=i^2i$, $mi^2=i^2m$, and $ji^2=i^2j$. Also $j^2t=tj^2$. Hence if $p=t+i^2$, then $jpj^{-1}=-t+i^2$, $zpz^{-1}=t+k^2$, $(jz)p(jz)^{-1}=t+k^2$. The field $F(t, i^2)$ is of order four since t is not a polynomial in i^2. The element p is of rank 4, since if it is of rank 2 then the resulting equation gives $2i^2$ equals a quantity in F, a contradiction. Hence the results here are the same as in the first case and we have an element p satisfying an equation with group G_4.

It remains to consider the cases where the algebra A is of type Γ. If A is of type E then, by the definition of algebras of type E, A contains the proper element and is of the desired type. If A is of type D, then by the definition of these cyclic algebras, A contains an element y which is a root of the quartic $w^4=d$, where d is in F. The group of this equation is either G_4 or G_8. If it is the group G_4 then y is the desired element. If not then the group of the equation is G_8 and not all of its roots are rational functions of one of them, since if its roots are x_1, $x_2=-x_1$, x_3, x_4, then if x_3 is a polynomial in x_1, by applying the substitution (34) of the group G_8, x_4 is the same rational function of x_1 and is equal to x_3, contrary to the hypothesis that the equation satisfied by y is irreducible. Hence we may apply the proof for algebras containing an element i satisfying (11) and obtain the desired result.

UNIVERSITY OF CHICAGO,
CHICAGO, ILL.

EXPANSIONS IN GENERALIZED APPELL POLYNOMIALS, AND A CLASS OF RELATED LINEAR FUNCTIONAL EQUATIONS*

BY

I. M. SHEFFER†

1. Introduction. In this paper we are primarily concerned with the solution of the linear differential equation of infinite order

$$\sum_{n=0}^{\infty} P_n(x) y^{(n)}(x) = f(x), \tag{1}$$

where the $P_n(x)$ are polynomials of bounded degree ($\leqq k$), and the relation of this equation to expansions in generalized Appell polynomials. Equation (1) is transformed into an equivalent contour integral equation, using the Laplace transformation; and this integral equation is then shown to lead to a solution which is itself a contour integral, with a kernel which satisfies a linear differential equation of order k. It is also shown that the contour integral equation is equivalent to an expansion question in generalized Appell polynomials.

In §2 we derive some simple properties of these polynomials. In §3 the equivalence between the above-mentioned differential and integral equations is shown, and the relation to the expansion problem developed. The resolving kernel for the general case (k) is introduced in §4, and the particular cases $k=0$, $k=1$, are treated in §§5, 6. The method of the Laplace transformation is then extended, in §§7, 8 respectively, to partial differential equations of infinite order, constant coefficients, and to Laurent differential equations, constant coefficients.‡

2. Generalized Appell polynomials. Let

$$A_i(t) \sim \sum_{n=0}^{\infty} \alpha_{in} t^n \qquad (i = 0, 1, \cdots, k) \tag{2}$$

---

* Presented to the Society, September 7, 1928; received by the editors in December, 1928.

† National Research Fellow.

‡ The attention of the writer has been called to two memoirs by S. Pincherle, the spirit of which is closely akin to that of the present paper: *Studi sopre alcune operazioni funzionali*, Memorie della Reale Accademia delle Scienze di Bologna, (4), vol. 7 (1886), pp. 393–442; *Sur la résolution de l'équation fonctionnelle* $\sum h_\nu \phi(x+\alpha_\nu) = f(x)$, Acta Mathematica, vol. 48 (1926), pp. 279–304.

be $(k+1)$ formal* power series, with $A_k(t) \not\equiv 0$, and define the *generalized Appell polynomials* $\{G_n(x)\}$, *of order* k, *by the (formal) expansion*†

$$e^{tx}\{A_0(t) + xA_1(t) + \cdots + x^kA_k(t)\} \sim \sum_{n=0}^{\infty} G_n(x)t^n. \tag{3}$$

If $A_k(0) \not\equiv 0$, $G_n(x)$ is a polynomial of degree $n+k$. If we multiply (5) through by e^{-tx}, and differentiate $(k+1)$ times with respect to x, we find that

$$0 \sim e^{-tx} \sum_{n=0}^{\infty} t^n \left\{G_n^{(k+1)}(x) - \binom{k+1}{1} G_{n-1}^{(k)}(x) + \cdots + (-1)^{k+1} G_{n-k-1}(x)\right\},$$

so that

LEMMA 1. *$G_n(x)$ satisfies the mixed relation*‡

$$G_n^{(k+1)}(x) - \binom{k+1}{1} G_{n-1}^{(k)}(x) + \binom{k+1}{2} G_{n-2}^{(k-1)}(x) - \cdots + (-1)^{k+1} G_{n-k-1}(x) = 0 \qquad (n = 0, 1, \cdots). \tag{4}$$

On expanding the left hand member of (3) in a power series in t we obtain an explicit form for $G_n(x)$:

$$G_n(x) = \sum_{i=0}^{k} x^i \left(a_{i0}\frac{x^n}{n!} + a_{i1}\frac{x^{n-1}}{(n-1)!} + \cdots + a_{in}\right). \tag{5}$$

Not only does (3) imply (4), but also conversely:

LEMMA 2. *Let $\{G_n(x)\}$, $n = 0, 1, \cdots$, be a set of functions satisfying (4) for $n = 0, 1, \cdots$. Then there exist constants a_{in} $(n = 0, 1, \cdots, \infty; i = 0, 1, \cdots, k)$ such that on setting (formally) $A_i(t) \sim \sum_0^\infty a_{in}t^n$, the relation (3) is formally satisfied.*

In virtue of (4) we have

$$0 \sim e^{-tx} \sum_{n=0}^{\infty} \left\{G_n^{(k+1)}(x) - \binom{k+1}{1} G_{n-1}^{(k)}(x) + \cdots + (-1)^{k+1} G_{n-k-1}(x)\right\}.$$

But the right hand member of this identity is the $(k+1)$st derivative with

---

* That is, the radius of convergence of $A_i(t)$ may be zero.

† The original Appell polynomials ($k=0$) were discussed by Appell, *Sur une classe de polynomes*, Annales Scientifiques de l'Ecole Normal Supérieure, (2), vol. 9 (1880), pp. 119–144. He has found for them some interesting properties.

‡ G's with negative subscripts are defined to be identically zero.

respect to x of $e^{-tz}\sum_0^\infty t^n G_n(x)$, so that this last expression is a polynomial in x of degree not exceeding k, with coefficients which are functions of t:*

$$e^{-tz}\sum_0^\infty t^n G_n(x) \sim A_0(t) + xA_1(t) + \cdots + x^k A_k(t).$$

It follows that the functions $G_n(x)$ are *polynomials*. To determine the functions $A_i(t)$, we differentiate the members of this last identity k times with respect to x, setting $x=0$ after each differentiation:

$$A_0(t) \sim \sum_0^\infty G_n(0)t^n,$$

$$A_1(t) \sim \frac{1}{1!}\sum_0^\infty \{G_n'(0) - G_{n-1}(0)\}t^n,$$

. . . . . .

$$A_k(t) \sim \frac{1}{k!}\sum_0^\infty \left\{G_n^{(k)}(0) - \binom{k}{1}G_{n-1}^{(k-1)}(0) + \binom{k}{2}G_{n-2}^{(k-2)}(0) - \cdots + (-1)^k\binom{k}{k}G_{n-k}(0)\right\}t^n.$$

This establishes the lemma.

COROLLARY. *A necessary and sufficient condition that a set of functions* $\{G_n(x)\}$, $n=0, 1, \cdots$, *be a set of generalized Appell polynomials of order* $\leqq k$, *is that* (4) *be satisfied,* $n=0, 1, 2, \cdots$.

In establishing the preceding results the question of convergence of the formal power series represented by $A_i(t)$ did not enter. However, in considering G_n-expansions of analytic functions, it is necessary to lay down restrictions. We make the following assumptions:

(A) The functions $A_i(t)$, $i=0, 1, \cdots, k$, are analytic, $|t|<R$.

(B) $A_k(t)$ has at most a finite number† of zeros in $|t|<R$.

(C) The analytic functions $f(x)$ whose expansions we consider are those of *exponential value* (*exp. val.*) less than R; i.e., those for which $\limsup_{n=\infty} |f^{(n)}(0)|^{1/n} < R$.

LEMMA 3. *If* $\{\lambda_n\}$ *is a sequence of numbers such that* $\limsup |\lambda_n|^{1/n} = 1/\rho < R$, *then* $\sum_0^\infty \lambda_n G_n(x)$ *converges uniformly in every bounded region, and defines a function of exponential value* $<R$.

* It is possible that $A_k(t)\equiv 0, A_{k-1}(t)\equiv 0, \cdots, A_{k-r}(t)\equiv 0, 0\leqq r\leqq k$.

† If $A_k(t)$ has infinitely many zeros in $|t|<R$, then we choose R' very close to, *and less than*, R, and consider the region $|t|<R'$. There will be only a finite number of zeros in $|t|<R'$.

Let r be any number $<R$, and let C be the circle $|t|=r$. Then in any bounded region $|x|\leqq X$,

$$|G_n(x)| = \left|\frac{1}{2\pi i}\int_C e^{tx}\{A_0(t)+\cdots+x^kA_k(t)\}\frac{dt}{t^{n+1}}\right|$$
$$\leqq e^{rX}\cdot N(1+X+\cdots+X^k)/r^n = M/r^n,$$

where N is the maximum of $|A_0(t)|,\cdots,|A_k(t)|$ on C. Choose r to satisfy $r>1/\rho$. Then in $|x|\leqq X$, $\sum_0^\infty|\lambda_nG_n(x)|\leqq M\sum_0^\infty|\lambda_n|/r^n$, which converges. The series represents, then, an entire function. Denote it by $g(x)$. Now $|a_{in}|\leqq M/(R-\epsilon)^n$, $i=0,\cdots,k$, so that

$$A_0(t)+\cdots+x^kA_k(t)\ll M(1+x+\cdots+x^k)\sum_0^\infty t^n(R-\epsilon)^{-n},$$

and

$$e^{tx}\{A_0(t)+\cdots+x^kA_k(t)\}\ll M(1+x+\cdots+x^k)\sum_0^\infty t^n[x^n/n!+\cdots+(R-\epsilon)^{-n}].$$

Hence

$$G_n(x)\ll M(1+x+\cdots+x^k)[x^n/n!+\cdots+(R-\epsilon)^{-n}].$$

Also, $|\lambda_n|\leqq N(\rho-\epsilon)^{-n}$; therefore

$$\sum_{n=0}^\infty\lambda_nG_n(x)\ll MN(1+x+\cdots+x^k)\sum_{n=0}^\infty(x^n/n!)(\rho-\epsilon)^{-n}\{\sum_{s=0}^\infty(R-\epsilon)^{-s}(\rho-\epsilon)^{-s}\}.$$

The series in the brace converges, since $R\rho>1$ and ϵ can be taken arbitrarily small; let its value be K. Then

$$g(x)=\sum_0^\infty\lambda_nG_n(x)\ll MNK(1+x+\cdots+x^k)e^{x/(\rho-\epsilon)}.$$

Now exp. val. of $e^{x/(\rho-\epsilon)}$ is $1/(\rho-\epsilon)$, $<R$ for ϵ small enough, and multiplying through by $(1+x+\cdots+x^k)$ does not affect the exponential value. Hence $g(x)$ is of exp. val. $<R$, as was to be proved.

3. **A differential equation of infinite order and a contour integral equation.** We return to the differential equation (1). Let us denote by $A[y(x)]$ the differential operator

$$A[y]\equiv a_0y(x)+\cdots+a_ny^{(n)}(x)+\cdots,$$

where $A(t)\sim\sum_0^\infty a_nt^n$. Then equation (1) can be expressed, by a suitable rearrangement of terms, as follows:

$$A_0[y]+xA_1[y]+\cdots+x^kA_k[y]=f(x). \tag{6}$$

We assume in what follows that conditions (A), (B), (C) *of the preceding section are fulfilled.*

Such equations have been considered by Perron,* Hilb, von Koch and others. We shall treat them from a new point of view. Let us look for solutions $y(x)$ which are of exp. val. $<R$. If $y(x)=\sum_0^\infty \lambda_n x^n/n!$ is of exp. val. $\mu<R$, and if $Y(x)=\sum_n^\infty \lambda_n x^n$, then $Y(x)$ has a radius of convergence $1/\mu$; and if C is a contour surrounding the origin and lying in $|t|<1/\mu$, then

$$y(x) = \frac{1}{2\pi i}\int_C \frac{Y(t)}{t} e^{x/t}dt. \tag{7}$$

Now it is readily established that if $A(t)$ is analytic in $|t|<R$, and $y(x)$ is of exp. val. $\mu<R$, then†

$$A[y] = \frac{1}{2\pi i}\int_C \frac{Y(t)}{t} A\left(\frac{1}{t}\right) e^{x/t}dt, \quad 1/R<|t|<1/\mu, \tag{8}$$

and $A[y]$ is of exp. val. $\leqq\mu$. That is, the operator A can be taken under the integral sign. It follows that (6) *is equivalent to the contour integral equation*

$$f(x) = \frac{1}{2\pi i}\int_C \frac{Y(t)}{t}\left\{A_0\left(\frac{1}{t}\right)+xA_1\left(\frac{1}{t}\right)+\cdots+x^kA_k\left(\frac{1}{t}\right)\right\} e^{x/t}dt, \tag{9}$$

the sense of the equivalence being as follows:

THEOREM 1. *If $y(x)=\sum_0^\infty \lambda_n x^n/n!$, of exp. val. $\mu<R$, is a solution of* (6) *then $Y(x)=\sum_0^\infty \lambda_n x^n$ is a solution of* (9), *C being a contour surrounding the origin and lying in $1/R<|t|<1/\mu$. Conversely, if $Y(x)$, with radius of convergence $1/\mu$, where $\mu<R$, is a solution of* (9), *then $y(x)$ is of exp. val. μ, and is a solution of* (6).‡

Let us assume in (9) the expansion

$$Y(t) = \sum_0^\infty \lambda_n t^n. \tag{10}$$

The term in $1/t$ in the integrand is the term independent of t in the Laurent expansion of $Y(t)\sum_0^\infty G_n(x)/t^n$, so that we have

$$f(x) = \sum_0^\infty \lambda_n G_n(x). \tag{11}$$

---

* For a list of references see Sheffer, *Linear differential equations of infinite order, with polynomial coefficients of degree one*, to appear in the Annals of Mathematics.

† $A[e^{\lambda x}]=A(\lambda)e^{\lambda x}$ for all $|\lambda|<R$.

‡ The proof is immediate.

There is, then, a very close relation between equations (6) and (9) on the one hand and G_n-expansions on the other. In precise form:

THEOREM 2. *Let $f(x)$ be of exp. val. $<R$. If $y(x)=\sum_0^\infty \lambda_n x^n/n!$, of exp. val. $<R$, is a solution of (6), then $f(x)$ possesses the expansion (11), which converges uniformly in every bounded region; and* $\limsup |\lambda_n|^{1/n}<R$. *Conversely, if* $\limsup |\lambda_n|^{1/n}<R$, *and $f(x)$ is defined by (11), then $f(x)$ and $y(x)=\sum_0^\infty \lambda_n x^n/n!$ are of exp. val. $<R$, and $y(x)$ is a solution of (6).*

The theorem follows at once on using (9) and Lemma 3.

DEFINITION. An expansion for $f(x)$: $f(x)=\sum_0^\infty \lambda_n G_n(x)$ will be termed a *proper expansion* (with respect to the given set $\{G_n(x)\}$) if $\limsup |\lambda_n|^{1/n}<R$.*

Theorem 2 relates to proper expansions, and may be (partially) restated as follows:

COROLLARY. *Let $f(x)$ be of exp. val. $<R$. Every proper expansion of $f(x)$ is equivalent to a solution (of exp. val. $<R$) of equation (6).*

A remarkable property of G_n-expansions is that they may permit of zero expansions; i.e., the function $f(x)\equiv 0$ may have one or more linearly independent expansions in which not all the coefficients vanish. This property is a consequence of the fact that the homogeneous equation corresponding to (6) may have solutions other than $y(x)\equiv 0$:

THEOREM 3.† *Every proper expansion of the function zero, in which not all the coefficients vanish, is equivalent to a solution $y(x)\not\equiv 0$, of exp. val. $<R$, of the homogeneous equation*

$$A_0[y] + xA_1[y] + \cdots + x^k A_k[y] = 0.$$

For equation (6) Perron‡ has obtained the following theorem:

Let m be the number of zeros (multiple zeros counted multiply) of $A_k(t)$ in $|t|<R$, and let s be the number of linearly independent solutions $D(t)$, which are analytic everywhere in $|t|<R$, of the equation

$$A_0(t)D(t) + A_1(t)D'(t) + \cdots + A_k(t)D^{(k)}(t) = 0. \tag{12}$$

Then the number of linearly independent solutions $y(x)$, of exp. val. $<R$, of the equation

---

* It follows by Lemma 3 that if $f(x)$ has a proper expansion, then $f(x)$ is of exp. val. $<R$.

† The proof follows that of Theorem 2, the expansion $0=\sum_0^\infty \lambda_n G_n(x)$ corresponding to the solution $y(x)=\sum_0^\infty \lambda_n x^n/n!$

‡ Perron, *Lineare Differentialgleichungen unendlich hoher Ordnung mit ganzen rationalen Koeffizienten*, Mathematische Annalen, vol. 84 (1921), pp. 31–42.

$$A_0[y] + xA_1[y] + \cdots + x^k A_k[y] = 0 \tag{13}$$

is precisely $m-k+s$; and the non-homogeneous equation (6) will possess a solution of exp. val. $<R$ for *all* functions $f(x)$ of exp. val. $<R$ if and only if (12) has no solution analytic everywhere in $|t|<R$.*

Applying this existence theorem and Theorem 2 (Corollary) and Theorem 3, we have

THEOREM 4. *A necessary and sufficient condition that a proper expansion exist for all functions $f(x)$ of exp. val. $<R$ is that equation* (12) *have no solution $D(t)\not\equiv 0$ which is everywhere analytic in $|t|<R$. The number of linearly independent† proper expansions of the function zero is precisely‡ $m-k+s$.*

4. **The resolving kernel.§** By Theorem 4, if (12) has a solution $D(t)$ analytic in $|t|<R$, not every function $f(x)$ of exp. val. $<R$ has a proper expansion. Let us then consider the contrary case: where every solution $D(t)$ has some singularity in $|t|<R$, so that solutions of (6) always exist.

We wish to invert the integral equation (9). Now (9) is in the form

$$f(x) = \frac{1}{2\pi i}\int_C \left\{ Y(t)\Theta\left(\frac{1}{t}\,;\, x\right)\right\}\frac{dt}{t}.$$

If then we interchange the rôles of f and Y (which is to introduce F and y), we are led to consider as a solution of the original equation (6) an expression of the form

$$y(x) = \frac{1}{2\pi i}\int_C \frac{F(t)}{t}\Lambda\left(\frac{1}{t}\,;\, x\right)dt, \tag{14}$$

---

* That not *all* functions $f(x)$ yield a solution of (6) in the case that (12) has a solution analytic everywhere in $|t|<R$, may be seen as follows:

If $A(t), B(t)$ are analytic in $|t|<R$, and $y(x)$ is of exp. val. $<R$, it is permissible to operate with A on $x^iB[y]$:

$$A[x^iB[y]] \equiv x^iAB[y] + \binom{i}{1}x^{i-1}A'B[y] + \cdots + \binom{i}{i}A^{(i)}B[y].$$

Here $A^{(i)}$ stands for the operator obtained by differentiating $A(t)=\sum_0^\infty a_n t^n$ i times, and $A^{(i)}B$ is the operator defined by the product $A^{(i)}(t)B(t)$, the multiplication of these power series being as usual.

Let $D(t)\not\equiv 0$ satisfy (12) and be analytic in $|t|<R$. Then on operating on (6) with D, as we may, we find that the term on the left which has no factor x, drops out in virtue of (12), so that the left hand member has a factor x. The right hand member is $D[f(x)]$, and this will not in general vanish at the origin. Consequently there will not always be a solution $y(x)$ of exp. val. $<R$.

† It is clear that linearly independent solutions yield linearly independent expansions.

‡ If $m-k+s$ is negative, zero will be understood.

§ We owe to Professor Tamarkin suggestions which have improved the presentation of this section, particularly in stressing the fact that the resolving kernel is not always unique.

where C is a contour surrounding $t=0$ and lying in $1/R<|t|<1/\lambda$, $f(x)$ being of exp. val. λ. The function $\Lambda(t;x)$ is to be determined.

On substituting (14) into (6) we arrive at the equation

$$f(x)=\frac{1}{2\pi i}\int_C \frac{F(t)}{t}\{A_0[\Lambda]+xA_1[\Lambda]+\cdots+x^kA_k[\Lambda]\}dt,$$

and this will certainly be true (see (7)) if the brace in the integrand reduces to e^{xt}. The function $\Lambda(t;x)$ is then to satisfy (with respect to x) the differential equation of infinite order

$$A_0[\Lambda(t;x)]+xA_1[\Lambda(t;x)]+\cdots+x^kA_k[\Lambda(t;x)]=e^{tx}. \tag{15}$$

DEFINITION. A function $\Lambda(t;x)$ which satisfies (15), is analytic in t for all t in $^*T$, and is (in x) of exp. val. $<R$ for all t in T, will be termed a *resolving kernel*, since it provides in (14) a solution of (6).

Consider the homogeneous equation

$$A_0[z(x)]+xA_1[z(x)]+\cdots+x^kA_k[z(x)]=0. \tag{16}$$

If $m-k+s>0$, this will be (by the Perron existence theorem) the number of linearly independent solutions $z_i(x)$ of (16), of exp. val. $<R$. It follows that there will not then be a unique resolving kernel:

LEMMA 4. *If there exists a resolving kernel, and $\Lambda(t;x)$ is such a one, then so is*

$$\text{(a)}\quad \Lambda(t;x)+\textstyle\sum_{i=1}^{r}\Upsilon_i(t)z_i(x),$$

if $r=m-k+s>0$; and this is the most general resolving kernel. Here the functions $\Upsilon_i(t)$ are subject only to the condition that they be analytic for t in T.

If $m-k+s\leqq 0$, (16) has no solution of exp. val. $<R$ other than $z(x)\equiv 0$. Hence

LEMMA 4′. *If $m-k+s\leqq 0$, a resolving kernel, if it exists, is unique.*

Let us denote by A_x the operator on the left of (15), so that $A_x[\Lambda(t;x)]=e^{tx}$; and let B_t be the operator

$$B_t[w(t)]\equiv A_0(t)w+A_1(t)w'+\cdots+A_k(t)w^{(k)}. \tag{17}$$

---

* T is the region $|t|<R$ with the zeros of $A_k(t)$ deleted. It is not strictly necessary to demand the analyticity and exp. val. conditions throughout T: it would suffice to consider only those values of t for which $1/t$ lies on C. But since the contour C may vary in T as we choose different functions $f(x)$ it is simpler to use the region T.

Then

$$A_x[e^{tx}] = B_t[e^{tx}] = e^{tx}\sum_{i=0}^{k} x^i A_i(t),$$

and*

$$A_x[B_t[\Lambda]] = B_t[A_x[\Lambda]] = B_t[e^{tx}] = A_x[e^{tx}], \tag{18}$$

so that

$$B_t[\Lambda(t;x)] = e^{tx} + \phi(t;x), \tag{19}$$

where $\phi(t;x)$ is a solution of $A_x[\phi]=0$. If Λ is a resolving kernel, then it is analytic in t (in the region T) and is (in x) of exp. val. $<R$ (t in T). The same is then true of $B_t[\Lambda]$, and hence of ϕ. That is,†

$$\phi(t;x) = \sum_{i=1}^{r} \Upsilon_i(t) z_i(x), \tag{b}$$

$\Upsilon_i(t)$ analytic in T.

LEMMA 5. *There is a solution $\Lambda_1(t;x)$ of the linear differential equation of order k*

$$B_t[\Lambda_1] \equiv A_0(t)\Lambda_1 + A_1(t)\frac{\partial \Lambda_1}{\partial t} + \cdots + A_k(t)\frac{\partial^k \Lambda_1}{\partial t^k} = e^{tx} + \phi(t;x), \tag{20}$$

which is analytic in t (in the region T) and which is (in x) of exp. val. $<R$ for t in T. Here ϕ is given by (b).†

The lemma follows from the properties of linear differential equations of finite order.

Consider such a function $\Lambda_1(t;x)$. We have

$$B_t[A_x[\Lambda_1]] = A_x[B_t[\Lambda_1]] = A_x[e^{tx}] = B_t[e^{tx}],$$

so that

$$A_x[\Lambda_1(t;x)] = e^{tx} + \psi(t;x), \tag{21}$$

where ψ is a solution of $B_t[\psi]=0$. The left hand member of (21) is of exp. val. $<R$ (see (8)). This must then also be true of $\psi(t;x)$.

Let $c_1(t), \cdots, c_k(t)$ be a set of linearly independent solutions of $B_t[c(t)]=0$. ($c_i(t)$ is clearly analytic in T.) Then we can write

---

* That $A_xB_t = B_tA_x$ is readily seen.

† If $A_x[z]=0$ has no solution (of exp. val. $<R$) other than $z(x)\equiv 0$, the sum in (b) drops out: $\sum_{i=1}^{r}\Upsilon_i(t)z_i(x)\equiv 0$.

$$\text{(c)} \qquad \psi(t;x) = \sum_{i=1}^{k} u_i(x)c_i(t),$$

where the $u_i(x)$ are functions of exp. val. $<R$.*

We come now to

THEOREM 5. *Let* (12) *have no solution* $D(t)\not\equiv 0$ *which is analytic everywhere in* $|t|<R$. *Then every function* $\Lambda(t;x)$ *defined by*

$$\text{(22)} \qquad \Lambda(t;x) = \Lambda_1(t;x) + \chi(t;x),$$

where χ *is an arbitrary solution of*†

$$\text{(23)} \qquad A_x[\chi(t;x)] = -\psi(t;x),$$

ψ *being given by* (c), *is a resolving kernel.*

Our hypothesis on equation (12) ensures us (by the Perron existence theorem) that a solution χ exists of the desired character.‡ Equation (22) then defines Λ to be analytic in t (in the region T) and of exp. val. $<R$ for all t in T. We need then only show that Λ satisfies (15). Now $A_x[\Lambda]=A_x[\Lambda_1]-\psi$, and this is, by (21), precisely e^{tz}.

Let us return to equation (15). We see (Lemmas 4 and 4′) that the resolving kernel just obtained is unique unless $m-k+s>0$, and in the case of non-uniqueness Λ may be augmented by $\sum_{i=1}^{r}\rho_i(t)z_i(x)$, the $\rho_i(t)$ being arbitrary functions analytic in T. This result is also reflected in

---

* For ψ certainly has the form (c) where the $u_i(x)$ are *some* functions of x. It remains to establish their exponential value. Now there exist k values of t in T: $t=t_1,\cdots,t_k$, such that the determinant Δ of the c_i's at these k points is not zero. For suppose the contrary, and let $t_1,\cdots,t_{k-1}$ be held fast, while t_k varies in T. Then Δ, which is a linear combination of $c_1(t_k),\cdots,c_k(t_k)$, must vanish identically (in t_k). Hence the coefficients of $c_1(t_k),\cdots,c_k(t_k)$ must vanish. This is true for all $t_1,\cdots,t_{k-1}$ in T, and as the coefficients in question are $(k-1)$-order minors of Δ, we have reduced the question to a $(k-1)$-order determinant. If now we permit t_{k-1} to vary in T, holding $t_1,\cdots,t_{k-2}$ fast, we reduce still further. Finally we arrive at a first-order determinant: $c_1(t_1)$, which must vanish identically in t_1. But this is manifestly a contradiction.

If we regard equations (c) (for $t=t_1,\cdots,t_k$) as a set of linear equations in $u_1(x),\cdots,u_k(x)$, we find (since $\Delta\not\equiv 0$) that $u_i(x)=$linear combination of $\psi(t_1;x),\cdots,\psi(t_k;x)$. Since $\psi(t;x)$ is of exp. val. $<R$ for all t in T, it follows then that $u_i(x)$ is of exp. val. $<R$.

† χ is also required to be analytic in t (in the region T) and (in x) of exp. val. $<R$ for all t in T.

‡ For the functions $u_i(x)$ of (c) are of exp. val. $<R$, so that functions $w_i(x)$, likewise of exp. val. $<R$, exist such that $A_x[w_i(x)]=u_i(x)$. Then the most general χ is given by

$$\chi(t;x) = -\sum_{i=1}^{k} c_i(t)w_i(x) + \sum_{i=1}^{r}\sigma_i(t)z_i(x),$$

where the $\sigma_i(t)$ are arbitrary functions which are analytic in T.

COROLLARY 1. *Any function* $\Lambda(t; x)$ *defined by* (22) *satisfies the linear differential equation of order k*

$$(24)\quad B_t[\Lambda] \equiv A_0(t)\Lambda + A_1(t)\frac{\partial \Lambda}{\partial t} + \cdots + A_k(t)\frac{\partial^k \Lambda}{\partial t^k} = e^{tz} + \sum_{i=1}^{r}\mu_i(t)z_i(x),$$

where the $\mu_i(t)$* *are suitably chosen functions analytic in* T, *and the* $z_i(x)$ *are a complete set*† *of linearly independent solutions of* $A_z[z(x)]=0$.

Given Λ defined by (22). Now $A_z[B_t[\chi]]=B_t[A_z[\chi]]=-B_t[\psi]=0$. Hence $B_t[\chi]=\sum_{i=1}^{r}\delta_i(t)z_i(x)$, $\delta_i(t)$ analytic in T. Also,

$$\phi(t;x)=\sum_{i=1}^{r}\Upsilon_i(t)z_i(x),$$

$\Upsilon_i(t)$ analytic in T. Formula (24) then follows from the relation

$$B_t[\Lambda] = B_t[\Lambda_1] + B_t[\chi] = e^{tz} + \{\phi(t; x) + B_t[\chi]\}.$$

COROLLARY 2. *There exists a resolving kernel* $\Lambda(t; x)$ *which satisfies*

$$(24')\qquad A_0(t)\Lambda + A_1(t)\frac{\partial \Lambda}{\partial t} + \cdots + A_k(t)\frac{\partial^k \Lambda}{\partial t^k} = e^{tz}.$$

The lemma asserts that a possible choice in (24) is $\mu_i(t)\equiv 0$. Two cases arise: (i) $m-k+s\leqq 0$. Then in (24) the sum $\sum_{i=1}^{r}\mu_i(t)z_i(x)$ drops out, and we have precisely (24′).

(ii) $m-k+s>0$. Then $\Lambda(t; x)$ is not unique. Let $\Lambda_2(t; x)$ satisfy (24), where all the $\mu_i(t)$ are zero. Then the general Λ is (by Lemma 4) $\Lambda=\Lambda_2+\sum_{i=1}^{r}\rho_i(t)z_i(x)$, $\rho_i(t)$ arbitrary (but analytic in T); and

$$B_t\left[\Lambda_2 + \sum_{i=1}^{r}\rho_i(t)z_i(x)\right] = e^{tz} + \sum_{i=1}^{r}z_i(x)\{\mu_i(t) + B_t[\rho_i(t)]\}.$$

On choosing $\rho_i(t)$ to satisfy $B_t[\rho_i(t)]=-\mu_i(t)$, the sum on the right drops out, and we have (24′).

That $\Lambda(t; x)$ can be augmented by the sum $\sum_{i=1}^{r}\rho_i(t)z_i(x)$ (when $m-k+s>0$) should imply (from (24)) that $B_t[\sum_{i=1}^{r}\rho_i(t)z_i(x)]$ is again of the form $\sum_{i=1}^{r}\beta_i(t)z_i(x)$, $\beta_i(t)$ analytic in T; and this is at once verified.

An immediate consequence of the definition of resolving kernel is

---

* The $\mu_i(t)$ will vary as we choose different Λ's.

† We recall that if $m-k+s\leqq 0$, the sum $\sum_{i=1}^{r}\mu_i(t)z_i(x)$ must drop out.

THEOREM 6.* *Let the same assumption on $D(t)$ be made as in Theorem* 5. *Then every resolving kernel $\Lambda(t; x)$ makes* (14) *a solution of the original equation* (6).

Returning to G_n-expansions we have

THEOREM 7. *Let $f(x)$ be of exp. val. $\lambda < R$ and let $\Lambda(t; x)$ be the resolving kernel given by Theorem* 5.† *Then $f(x)$ possesses the proper expansion*

$$f(x) = \sum_0^\infty \lambda_n G_n(x) \tag{25}$$

with

$$\lambda_n = \frac{1}{2\pi i}\int_C \frac{F(t)}{t}\Omega_n\left(\frac{1}{t}\right)dt, \tag{26}$$

where

$$\Lambda(t\,;x) = \sum_{n=0}^\infty \frac{\Omega_n(t)x^n}{n!}\cdot \tag{27}$$

Defining $\Omega_n(t)$ by (27) and λ_n by (26), we have, by (14), $y(x) = \sum_0^\infty \lambda_n x^n/n!$. The theorem now follows by use of Theorems 2 and 6.

The investigation of the resolving kernel $\Lambda(t; x)$, when it exists, is better effected through equation (24′) than through (15), since the theory of linear differential equations of finite order is well known. Yet even when dealing with (24′) it is difficult to obtain for the solutions a form which is convenient to handle. We shall content ourselves with a study of the cases $k=0$, $k=1$.

5. **The case $k=0$.** Here (15), (24′) reduce to

$$A_0[\Lambda] = e^{tx}, \qquad A_0(t)\Lambda(t; x) = e^{tx}, \tag{15a, 24'a}$$

and it is seen that

$$\Lambda(t\,;x) = e^{tx}A_0(t) = \sum_{n=0}^\infty \frac{t^n}{A_0(t)}\cdot\frac{x^n}{n!} \tag{27a}$$

---

* It is assumed that $f(x)$ is of exp. val. $\lambda < R$, and that the contour C of (14) surrounds the origin and lies in $1/R < |t| < 1/\lambda$. We naturally choose C so as not to pass through any point $1/\xi$, where ξ is a zero of $A_k(t)$, for at such a point $\Lambda(1/t; x)$ may cease to be analytic. It may be remarked that the formal step of taking the operators $A_0, xA_1, \cdots, x^kA_k$ under the integral sign in passing from (14) to (15) is easily justified.

† The same assumption on $D(t)$ is made as in Theorem 5.

satisfies both these equations. (27a) is then a resolving kernel* for equation (6a):

$$A_0[y] = f(x). \tag{6a}$$

On applying Theorems 6 and 7 we have

THEOREM 8. *If $f(x)$ is of exp. val. $\lambda<R$, then* (6a) *has a solution $y(x)$, of exp. val. $<R$, given by*

$$y(x) = \frac{1}{2\pi i}\int_C \frac{F(t)}{t}\,\frac{e^{x/t}dt}{A_0(1/t)}, \tag{14a}$$

C being a contour surrounding the origin and lying in $1/R<|t|<1/\lambda$; and $f(x)$ possesses the proper expansion

$$f(x) = \sum_0^\infty \lambda_n G_n(x) \tag{25a}$$

where†

$$\lambda_n = \frac{1}{2\pi i}\int_C \frac{F(t)}{t^{n+1}}\,\frac{dt}{A_0(1/t)}. \tag{26a}$$

It is possible to obtain solutions of the homogeneous equation $A_0[y]=0$ (in the case that $A_0(t)$ has at least one zero in $|t|<R$), as contour integrals. If ξ is a zero of $A_0(t)$ (in $|t|<R$), and p is the order, then there are p linearly independent solutions corresponding to this zero: $e^{\xi x}$, $xe^{\xi x}$, $\cdots$, $x^{p-1}e^{\xi x}$; and these give rise to the following p proper expansions of the function zero:‡

$$0 = \sum_0^\infty \lambda_n G_u(x),$$

where $\lambda_n=\xi^n$, $n\xi^{n-1}$, $n(n-1)\xi^{n-2}$, $\cdots$, $n(n-1)\cdots(n-p+2)\xi^{n-p+1}$.

6. Case $k=1$. For $k=1$ equations (15), (24′) become (15b), $A_0[\Lambda]+xA_1[\Lambda]=e^{tx}$, (24′b), $A_0(t)\Lambda+A_1(t)\partial\Lambda/\partial t=e^{tx}$.

A solution of (24′b) is furnished by

$$y(t\,;x) = e^{-\int(A_0/A_1)\,dt}\int \frac{e^{tx+\int(A_0/A_1)\,dt}}{A_1}\,dt,$$

and we have

$$A_0[y] + xA_1[y] = e^{-\int(A_0/A_1)\,dt}\int \frac{d}{dt}\left(e^{tx+\int(A_0/A_1)\,dt}\right)dt.$$

* We remark that (12) reduces to $A_0(t)D(t)=0$, which has only the solution $D(t)=0$.

† The case $k=0$ is treated and solved in Pincherle (Bologna Paper), loc. cit., pp. 423–424.

‡ We apply Theorem 3.

We are assuming that (12) (for $k=1$) has no solution $D(t)$ which is analytic throughout $|t|<R$. Therefore, at some zero $t=\xi$ of $A_1(t)$, in $|t|<R$, $D(t)$ fails to be analytic. Let $t=\xi$ be a zero of order s, and set

$$\text{(a)} \qquad \frac{A_0(t)(t-\xi)^s}{A_1(t)} = \sum_{n=0}^{\infty} l_n(t-\xi)^n,$$

so that

$$\text{(b)} \qquad l_{s-1} = \frac{1}{(s-1)!}\frac{d^{s-1}}{dt^{s-1}}\left\{\frac{A_0(t)(t-\xi)^s}{A_1(t)}\right\}_{t=\xi};$$

$$\text{(c)} \qquad B(t) = \frac{A_0(t)(t-\xi)^s}{A_1(t)} - l_{s-1}(t-\xi)^{s-1},$$

$$\text{(d)} \qquad e^{\int B(t)(t-\xi)^{-s}dt} = \sum_{n=-\infty}^{\infty} s_n(t-\xi)^n,$$

$$\text{(e)} \qquad e^{(t-\xi)x+\int B(t)(t-\xi)^{-s}dt} = \sum_{n=-\infty}^{\infty} r_n(x)(t-\xi)^n,$$

so that

$$\text{(f)} \qquad r_n(x) = s_n + \frac{x}{1!}s_{n-1} + \frac{x^2}{2!}s_{n-2} + \cdots \qquad (n = 0, \pm 1, \pm 2, \cdots).$$

Then

$$\text{(g)} \qquad e^{tx+\int (A_0/A_1)dt} = e^{\xi x}(t-\xi)^{l_{s-1}}\sum_{n=-\infty}^{\infty} r_n(x)(t-\xi)^n,$$

and

$$\text{(h)} \qquad \frac{d}{dt}(e^{tx+\int (A_0/A_1)dt}) = e^{\xi x}(t-\xi)^{l_{s-1}-1}\sum_{n=-\infty}^{\infty}(l_{s-1}+n)r_n(x)(t-\xi)^n.$$

Let us integrate (h), *introducing no "constant" of integration.* Two cases arise:

(α) l_{s-1} **is not an integer (positive, negative, or zero).** Then (h) integrates back into (g) and we have $A_0[y]+xA_1[y]=e^{tx}$. Therefore

$$\text{(i)} \qquad \Lambda(t\,;x) = e^{-\int [A_0(t)/A_1(t)]dt}\int \frac{e^{tx+\int [A_0(t)/A_1(t)]dt}}{A_1(t)}dt$$

is a resolving kernel. Here the integration is understood to be performed by expanding the integrand in a series about $t=\xi$, and introducting no "constant" of integration.

(β) l_{s-1} **is an integer.** Then the Laurent expansion (g) has a term

$e^{\xi x}r_{-l_{s-1}}(x)$ independent of t. Therefore on integrating (h) (and introducing no "constant" of integration),

$$\int \frac{d}{dt}(e^{tx+\int(A_0/A_1)\,dt})dt = e^{tx+\int(A_0/A_1)\,dt} - e^{\xi x}r_{-l_{s-1}}(x).$$

Let $u(x)$ be a solution* (of exp. val. $<R$) of

$$A_0[u] + xA_1[u] = e^{\xi x}r_{-l_{s-1}}(x). \tag{k}$$

Then

$$\Lambda(t\,;x) = e^{-\int[A_0(t)/A_1(t)]\,dt}\left[u(x) + \int \frac{e^{tx+\int[A_0(t)/A_1(t)]\,dt}}{A_1(t)}dt\right] \tag{l}$$

is a resolving kernel. Integration is as in (i).

To sum up:

THEOREM 9. *Let equation* (12) (*for* $k=1$) *have no solution* $D(t)$ *which is everywhere analytic in* $|t|<R$, *so that at some zero* $t=\xi$ *of* $A_1(t)$ (*in* $|t|<R$) $D(t)$ *is not analytic; and let the order of* $t=\xi$ *be* s. *Then a resolving kernel* $\Lambda(t;x)$ *for equation*

$$A_0[y] + xA_1[y] = f(x) \tag{6b}$$

is given by (i) *or* (l), *according as* l_{s-1} *is not or is an integer* (*positive, negative, or zero*). *According to the case,* $f(x)$ (*of exp. val.* $<R$) *possesses the proper expansion*

$$f(x) = \sum_{n=0}^{\infty}\left\{\frac{1}{2\pi i}\int_C \frac{F(t)}{t}\Theta_n\left(\frac{1}{t}\right)dt\right\}G_n(x), \tag{26b$_1$}$$

$$f(x) = \sum_{n=0}^{\infty}\left\{\frac{1}{2\pi i}\int_C \frac{F(t)}{t}\left[\Theta_n\left(\frac{1}{t}\right) + \Delta_n\left(\frac{1}{t}\right)dt\right\}G_n(x), \tag{26b$_2$}$$

where

$$\Theta_n(t) = e^{-\int[A_0(t)/A_1(t)]\,dt}\int \frac{t^n e^{\int[A_0(t)/A_1(t)]\,dt}}{A_1(t)}dt, \quad \Delta_n(t) = u_n e^{-\int[A_0(t)/A_1(t)]\,dt},$$

and where u_n *is defined by* $u(x)=\sum_0^\infty u_n x^n/n!$. *And in either case a solution of* (6b) *is given by*

$$y(x) = \frac{1}{2\pi i}\int_C \frac{F(t)}{t}\Lambda\left(\frac{1}{t}\,;x\right)dt. \tag{14b}$$

---

* The right member of (k) is seen to be of exp. val. $<R$, so that by our hypothesis on $D(t)$, a solution $u(x)$ exists.

7. **Partial differential equations of infinite order, constant coefficients.** The equivalence between linear differential equations of infinite order and contour integral equations, such as has been established in the preceding sections, can be extended to other equations. In this section and the following one we consider two such extensions.

Denote by $A_{xy}[z(x, y)]$ the partial differential operator in two* independent variables

$$A_{xy}[z] \equiv \sum_{m,n=0}^{\infty} a_{mn} \frac{\partial^{m+n} z}{\partial x^m \partial y^n}, \tag{28}$$

where

$$A(t_1, t_2) = \sum_{m,n=0}^{\infty} a_{mn} t_1^m t_2^n .$$

DEFINITION. If to the function $z(x; y)$, entire in x and y, there correspond two positive numbers λ_1, λ_2 such that for every $\epsilon > 0$,

$$\left| \frac{\partial^{m+n} z(0, 0)}{\partial x^m \partial y^n} \right| < M(\epsilon)(\lambda_1 + \epsilon)^m (\lambda_2 + \epsilon)^n,$$

then $z(x, y)$ will be said to be *of finite exponential value*; and if λ_1, λ_2 are the smallest numbers† for which this inequality holds, then $z(x, y)$ is *of exp. val.* (λ_1, λ_2).

LEMMA 6. *If $A(t_1, t_2)$ is analytic in $|t_1| < R_1$, $|t_2| < R_2$, and $z(x, y)$ is of exp. val. (λ_1, λ_2) with $\lambda_1 < R_1$, $\lambda_2 < R_2$, then $A_{xy}[z]$ exists and is of exp. val. (μ_1, μ_2) where $\mu_1 \leqq \lambda_1$, $\mu_2 \leqq \lambda_2$.*

By hypothesis we have $|a_{mn}| < N(\epsilon)(R_1 - \epsilon)^{-m}(R_2 - \epsilon)^{-n}$, where

$$A(t_1, t_2) = \sum a_{mn} t_1^m t_2^n ;$$

and

$$z(x, y) \ll M(\epsilon) \sum (\lambda_1 + \epsilon)^m (\lambda_2 + \epsilon)^n x^m y^n / (m!n!) = M(\epsilon) e^{(\lambda_1+\epsilon)x + (\lambda_2+\epsilon)y}.$$

Therefore

$$A_{xy}[z] \ll M(\epsilon) N(\epsilon) e^{(\lambda_1+\epsilon)x + (\lambda_2+\epsilon)y} \sum_{m,n=0}^{\infty} \left(\frac{\lambda_1 + \epsilon}{R_1 - \epsilon} \right)^m \left(\frac{\lambda_2 + \epsilon}{R_2 - \epsilon} \right)^n,$$

and for ϵ sufficiently small the series on the right converges. The lemma now follows from the fact that e^{rx+sy} is of exp. val. $(|r|, |s|)$, and from the fact that $\epsilon > 0$ is arbitrary.

* The extension of the theory to k independent variables is immediate.

† If z is of finite exp. val. then two "smallest numbers" exist.

LEMMA 7. *If $A(t_1, t_2)$, $B(t_1, t_2)$ are analytic in $|t_1|<R_1$, $|t_2|<R_2$, and $z(x, y)$ is of exp. val. $(\lambda_1, \lambda_2)<(R_1, R_2)$,* then*

$$B_{xy}[A_{xy}[z]] = A_{xy}[B_{xy}[z]] = C_{xy}[z],$$

where $C(t_1, t_2)=A(t_1, t_2)B(t_1, t_2)$.

The identities are seen to be formally true, and the convergence of all the series involved can be established by the method of Lemma 6.

LEMMA 8. *Let $z(x, y)=\sum\lambda_{mn}x^m y^n/(m!n!)$ be of exp. val. $(\lambda_1, \lambda_2)<(R_1, R_2)$, and set $Z(x, y)=\sum\lambda_{mn}x^m y^n$. Then $Z(x, y)$ is analytic in $|x|<1/\lambda_1$, $|y|<1/\lambda_2$, and*

$$z(x, y) = \left(\frac{1}{2\pi i}\right)^2 \int_{C_1}\int_{C_2} \frac{Z(t_1, t_2)}{t_1 t_2} e^{x/t_1+y/t_2}dt_1dt_2, \tag{29}$$

C_1, C_2 being contours in the t_1, t_2 planes respectively, surrounding $t_1=0$, $t_2=0$ and lying in $1/R_1<|t_1|<1/\lambda_1$, $1/R_2<|t_2|<1/\lambda_2$.

That $Z(x, y)$ is analytic in the region given follows at once. The integrand of (29) is then analytic on C_1, C_2, so that we may integrate, say first with respect to t_1, then t_2. Formula (29) results if we use relation (7) of §3.

If $(\lambda_1, \lambda_2)<(R_1, R_2)$ it is permissible to operate with A_{xy} under the integral sign, so that we have

THEOREM 10. *If $A(t_1, t_2)$ is analytic in $|t_1|<R_1$, $|t_2|<R_2$, and $f(x, y)$ is of exp. val. $(\lambda_1, \lambda_2)<(R_1, R_2)$, then the partial differential equation*

$$A_{xy}[z] = f(x, y) \tag{30}$$

is equivalent to the contour multiple integral equation†

$$f(x, y) = \left(\frac{1}{2\pi i}\right)^2 \int_{C_1}\int_{C_2} \frac{Z(t_1, t_2)}{t_1 t_2} A\left(\frac{1}{t_1}, \frac{1}{t_2}\right) e^{x/t_1+y/t_2}dt_1dt_2, \tag{31}$$

C_1, C_2 being the contours of (29).

The sense of the equivalence is as in Theorem 1: To every solution $z(x, y)=\sum\lambda_{mn}x^m y^n/(m!n!)$ of (30), of exp. val. $(m_1, m_2)<(R_1, R_2)$, corresponds a solution $Z(x, y)=\sum\lambda_{mn}x^m y^n$ of (31), analytic in $|t_1|<1/m_1$, $|t_2|<1/m_2$; and conversely. And each such solution is equivalent to a *proper expansion*‡

* The notation $(\lambda_1, \lambda_2)<(R_1, R_2)$ means that $\lambda_1<R_1$, $\lambda_2<R_2$.

† We use the identity $A_{xy}[e^{rx+sy}]=A(r, s)e^{rx+sy}$, $(|r|, |s|)<(R_1, R_2)$.

‡ For two variables a proper expansion is defined as one for which $\epsilon>0$ exists such that $|\lambda_{mn}|<M(\epsilon)(R_1-\epsilon)^m(R_2-\epsilon)^n$.

$$f(x, y) = \sum_{m,n=0}^{\infty} \lambda_{mn} G_{mn}(x, y) \tag{32}$$

where $\{G_{mn}(x, y)\}$ is the set of Appell polynomials in two variables defined by

$$e^{t_1 x + t_2 y} A(t_1, t_2) = \sum_{m,n=0}^{\infty} G_{mn}(x, y) t_1^m t_2^n . \tag{33}$$

As in the case of one independent variable, a solution of (30) can be obtained as a contour integral:

THEOREM 11.* *If $A(t_1, t_2)$ is analytic in $|t_1| < R_1$, $|t_2| < R_2$, and $f(x, y)$ is of exp. val. $(\lambda_1, \lambda_2) < (R_1, R_2)$, then a solution $z(x, y)$ of (30), of exp. val. $< (R_1, R_2)$, is given by*

$$z(x, y) = \left(\frac{1}{2\pi i}\right)^2 \int_{C_1} \int_{C_2} \frac{F(t_1, t_2)}{t_1 t_2} \cdot \frac{e^{x/t_1 + y/t_2}}{A\left(\frac{1}{t_1}, \frac{1}{t_2}\right)} dt_1 dt_2 . \tag{34}$$

Here C_1, C_2 are the contours of (31) and (29), and $F(t_1, t_2)$ is the function related to $f(t_1, t_2)$ in the same way that $Z(x, y)$ is related to $z(x, y)$.

COROLLARY. *If $f(x, y)$ is of exp. val. $(\lambda_1, \lambda_2) < (R_1, R_2)$ it possesses the proper expansion*

$$f(x, y) = \sum_{m,n=0}^{\infty} \lambda_{mn} G_{mn}(x, y) \tag{35}$$

where

$$\lambda_{mn} = \left(\frac{1}{2\pi i}\right)^2 \int_{C_1} \int_{C_2} \frac{F(t_1, t_2)}{t_1^{m+1} t_2^{n+1}} \frac{dt_1 dt_2}{A(1/t_1, 1/t_2)} . \tag{36}$$

It is clear that the results of this section carry over at once to any number of independent variables.

8. **"Laurent" differential equations of infinite order, constant coefficients.** Let $A(t) \sim \sum_{-\infty}^{\infty} a_n t^n$ be a formal Laurent series, and denote by $A_L[y(x)]$ the "Laurent" differential operator

$$A_L[y] = \sum_{n=-\infty}^{\infty} a_n y^{(n)}(x),$$

where

$$y^{(-k)}(x) = \int_0^x \cdots \int_0^x y(x) dx \cdots dx, \quad k > 0.$$

* The proof is immediate, using the relation (29) with f, F in place of z, Z.

Then we term

$$A_L[y] = f(x) \tag{37}$$

a "*Laurent*" *differential equation of infinite order.*

We make the following assumptions:

(a) $A(t) = \sum_{-\infty}^{\infty} a_n t^n$ converges, $0 < |t| < R$;

(b) $f(x)$ is of exponential value less than R.

LEMMA 9. *We have*

$$\int_0^x \cdots \int_0^x e^{tx} dx \cdots dx = \frac{1}{t^n} e^{tx} - \left\{ \frac{x^{n-1}}{(n-1)!} \frac{1}{t} + \frac{x^{n-2}}{(n-2)!} \frac{1}{t^2} + \cdots + \frac{x}{1!} \frac{1}{t^{n-1}} + \frac{1}{t^n} \right\}. \tag{38}$$

(n-fold)

LEMMA 10. *For every t in $0 < |t| < R$,*

$$A_L[e^{tx}] = A(t)e^{tx} - \Theta(t\,;\,x) \tag{39}$$

where

$$\Theta(t\,;\,x) = \sum_{n=1}^{\infty} \frac{a_{-n}}{t^n} \left\{ \frac{(tx)^{n-1}}{(n-1)!} + \cdots + \frac{tx}{1!} + 1 \right\}; \tag{40}$$

and $\Theta(t; x)$ is analytic for every x and for every t in $|t| > 0$.

(39) is a consequence of (38). We need then only consider $\Theta(t; x)$. Suppose $0 < |t|$. Then $1 + tx/1! + \cdots + (tx)^{n-1}/(n-1)! \ll e^{|x|t}$, so that $\Theta(t; x) \ll e^{|x|t} \sum_1^{\infty} |a_{-n}|/t^n$, and since this last series converges for all $|t| > 0$, Θ is analytic for all $|t| > 0$ and for all x.

On using (7) of §3, and (39), we can express (37) as a *contour integral equation*:

$$f(x) = \frac{1}{2\pi i} \int_C \frac{Y(t)}{t} \left\{ A\left(\frac{1}{t}\right) e^{x/t} - \Theta\left(\frac{1}{t}\,;\,x\right) \right\} dt\,; \tag{41}$$

and (37), (41) are equivalent in the sense given by Theorem 1 for equations (6), (9).

Equation (41) can however be simplified. For from the definition of Θ we find that

LEMMA 11. *$\Theta(1/t; x)$ is analytic in t at $t = 0$, and vanishes there.*

We have

$$\Theta\left(\frac{1}{t}\,;\,x\right) = t \sum_{n=1}^{\infty} a_{-n} \left\{ \frac{x^{n-1}}{(n-1)!} + \frac{x^{n-2}}{(n-2)!} t + \cdots + \frac{x}{1!} t^{n-2} + t^{n-1} \right\}.$$

Now $|a_{-n}| \leqq M$, so that $\Theta(1/t;\ x) \ll Mt\{e^{|x|}+te^{|x|}+t^2e^{|x|}+\cdots\} = Mte^{|x|} \cdot (1-t)^{-1}$; and this proves the lemma.

COROLLARY. *$\Theta(1/t; x)$ is an entire function in both x and t.*

Since $\Theta(1/t;\ x) = 0$ for $t = 0$, it follows that

$$\frac{1}{2\pi i}\int_C \left\{\frac{Y(t)}{t}\right\}\Theta\left(\frac{1}{t}\ ;\ x\right)dt = 0,$$

if $Y(t)$ is analytic in and on C. Consequently we have

THEOREM 12. *The "Laurent" equation* (37) *is equivalent (in the sense of Theorem* 1) *to the contour integral equation*

$$f(x) = \frac{1}{2\pi i}\int_C \frac{Y(t)}{t}A\left(\frac{1}{t}\right)e^{x/t}dt. \tag{42}$$

THEOREM 13. *If $y(x) = \sum_0^\infty \lambda_n x^n/n!$, of exp. val. $< R$, is a solution of* (37), *then $f(x)$ possesses the proper* Appell function expansion*

$$f(x) = \sum_0^\infty \lambda_n G_n(x)\ ; \tag{43}$$

and conversely.† Here $G_n(x)$ is defined by

$$e^{tx}A(t) = \sum_{n=-\infty}^{\infty} G_n(x)t^n. \tag{44}$$

---

* That is, $\limsup |\lambda_n|^{1/n} < R$.

† The proof is immediate.

UNIVERSITY OF CHICAGO,
CHICAGO, ILL.

SYSTEMS OF INFINITELY MANY LINEAR DIFFERENTIAL EQUATIONS OF INFINITE ORDER, WITH CONSTANT COEFFICIENTS*

BY

I. M. SHEFFER†

1. INTRODUCTION

The writer has given existence theorems governing a finite system of linear differential equations of infinite order, with constant coefficients.‡ There, the method of attack was one of operators, and by means of suitable differential operators the system was first reduced to a single equation of infinite order, after which certain solutions of this one equation were shown to be solutions of the original system.

Let us adopt the following notation for a differential operator:

$$A[y] \equiv a_0 y(x) + a_1 y'(x) + \cdots + a_n y^{(n)}(x) + \cdots,$$

where the a_n's are constants. The system of equations of which we speak has then the form (homogeneous and non-homogeneous)

$$\text{(I)} \qquad A_{i1}[y_1] + A_{i2}[y_2] + \cdots + A_{in}[y_n] = 0 \qquad (i = 1,2, \cdots, n);$$

$$\text{(II)} \qquad A_{i1}[y_1] + A_{i2}[y_2] + \cdots + A_{in}[y_n] = f_i(x) \qquad (i = 1,2, \cdots, n).$$

There are n equations in the n unknown functions $y_i(x)$.

Our assumptions, on the basis of which the existence theorems follow, are these:

(i) The $A_{ij}(t)$ are analytic§, $|t| < R$.

$$\text{(A) (ii)} \qquad \Delta(t) = \begin{vmatrix} A_{11}(t) \cdots A_{1n}(t) \\ \cdots\cdots\cdots\cdots \\ A_{n1}(t) \cdots A_{nn}(t) \end{vmatrix} \not\equiv 0.$$

(iii) The $f_i(x)$ are of *exponential value* $< R$. (For short: of exp. val. $< R$.) And the solutions sought are those functions $y_i(x)$ of exp. val. $< R$. We *define* exponential value as follows:

---

* Presented to the Society, September 7, 1928; received by the editors in December, 1928.

† National Research Fellow.

‡ Sheffer, *Systems of linear differential equations of infinite order, with constant coefficients*, to appear in the Annals of Mathematics. We shall refer to this as Annals Paper I.

§ That is, $A_{ij}(t) = \sum_{n=0}^{\infty} a_{ij,n} t^n$ converges, $|t| < R$.

$f(x)$ is of exp. val. λ if $f(x)$ is entire and $\limsup_{n=\infty} |f^{(n)}(0)|^{1/n} = \lambda$.

In the present paper we consider the extension to systems of infinitely many equations, and show how such systems can be thrown into the form of systems of contour integral equations in the complex plane. These integral equations are then inverted: we obtain solutions which are themselves contour integrals.* The method employs the Laplace transformation.†

In §2 we take up the non-homogeneous equation for an infinite system, and in §3 we consider the homogeneous case, where solutions are also obtained as contour integrals.

2. The non-homogeneous case, infinite system

We consider a system of infinitely many equations in infinitely many unknown functions:‡

$$\text{(III)}\quad A_{i1}[y_1] + A_{i2}[y_2] + \cdots + A_{in}[y_n] + \cdots = f_i(x) \quad (i = 1, 2, \cdots, \infty).$$

We assume that the following conditions are fulfilled:

(i) $A_{ij}(u)$ is analytic, $|u| < R$;

(B) (ii) $f_i(x)$ is of exp. val. $\lambda_i < \lambda < R$;

(iii) $|F_i(u)| \leqq M$ uniformly in i and in u, where u ranges over $|u| \leqq 1/\lambda$;

and

$$\text{(C)}\quad \text{(i)}\qquad \Delta(u) = \begin{vmatrix} A_{11} \cdots A_{1n} \cdots \\ A_{21} \cdots A_{2n} \cdots \\ \cdot \quad\quad \cdot \\ \cdot \quad\quad \cdot \\ \cdot \quad\quad \cdot \end{vmatrix}$$

is of *normal form* for every u in $|u| < R$; i.e.,

---

* The finite case is of course a particular case.

† The Laplace transformation has been used in the complex domain by S. Pincherle, *Sur la résolution de l'équation fonctionnelle* $\sum h_\nu \phi(x+\alpha_\nu) = f(x)$, Acta Mathematica, vol. 48 (1926), pp. 279–304; and in the domain of reals by F. Bernstein and G. Doetsch in a series of papers, of which we cite only two: F. Bernstein and G. Doetsch, *Probleme aus der Theorie der Wärmeleitung*, Mathematische Zeitschrift, vol. 22 (1925), pp. 285–292; G. Doetsch, *Die Integrodifferentialgleichungen vom Faltungstypus*, Mathematische Annalen, vol. 89 (1923), pp. 192–207.

‡ We can equally well consider the system

$$\sum_{j=-\infty}^{\infty} A_{ij}[y_j] = f_i(x) \qquad (i = -\infty, \cdots, \infty),$$

putting on the functions $A_{ij}(u)$, $f_i(x)$ the same conditions as we shall lay down for system (III). We arrive at a like existence theorem. For the properties of two-way infinite determinants see von Koch, *Sur les déterminants infinis et différentielles linéaires*, Acta Mathematica, vol. 16, pp. 217–249.

$\prod_{i=1}^{\infty} |A_{ii}(u)|$ converges and $\sum_{i,j=1}^{\infty}{}_{i\neq j} |A_{ij}(u)|$ converges;*

(ii) the convergence of the above product and series is uniform in every closed region lying in $|u|<R$;

(iii) $\Delta(u)\not\equiv 0$.

To discuss system (III) we require some preliminary lemmas.

LEMMA 1.† *If* $f(x)=\sum_0^\infty f_n x^n$ *is of exp. val.* λ, *and* $F(x)=\sum_0^\infty F_n x^n$ *where* $F_n=n!f_n$, *then* $F(x)$ *is analytic in* $|x|<1/\lambda$, *and*

$$f(x)=\frac{1}{2\pi i}\int_C \frac{F(u)}{u}e^{x/u}du, \tag{1}$$

where C is a contour surrounding the origin and lying in $|u|<1/\lambda$. *Conversely, if* $F(x)$ *has the radius of convergence* $1/\lambda$, *and* $f(x)$ *is given by* (1), *then* $f(x)$ *is of exp. val.* λ.

Let us agree, in what follows, to denote by a capital-letter function that function which is related to the corresponding small-letter function as $F(x)$ is to $f(x)$ in Lemma 1.

LEMMA 2.‡ *If* $A(t)$ *is analytic,* $|t|<R$, *and* $f(x)$ *is of exp. val.* $\lambda<R$, *then* $A[f]$ *is entire and is of exp. val.* $\leqq\lambda$.

LEMMA 3.§ *If* $A(t)$ *is analytic,* $|t|<R$, *and* ξ *is in* $|\xi|<R$, *then*

$$A[e^{tz}]=A(\xi)e^{\xi z}. \tag{2}$$

LEMMA 4. *If* $A(t)$ *is analytic,* $|t|<R$, *and* $y(x)$ *is of exp. val.* $\lambda<R$, *then*

$$A[y]=\frac{1}{2\pi i}\int_C \frac{Y(u)}{u}A\left(\frac{1}{u}\right)e^{x/u}du, \tag{3}$$

where C is a contour surrounding the origin and lying in $1/R<|u|<1/\lambda$.

We see that (3) follows from (1) formally if we use (2). It remains to justify taking the operator A under the integral sign. Now

---

* It follows that $\Delta(u)$ exists for all u in $|u|<R$. See F. Riesz, *Les Systèmes d'Équations Linéaires à une Infinité d'Inconnues* (Borel monograph), pp. 24–25.

† Sheffer, *Linear differential equations of infinite order, with polynomial coefficients of degree one,* to appear in the Annals of Mathematics.

‡ Sheffer, Annals Paper I.

§ Sheffer, Annals Paper I.

$$A[y] = \sum_0^\infty a_n y^{(n)}(x) = \sum_0^\infty a_n \cdot \frac{1}{2\pi i}\int_C \left\{\frac{Y(u)e^{x/u}}{u^{n+1}}\right\} du,$$

and since $\sum_0^\infty a_n/u^n$ converges uniformly to $A(1/u)$ on C, term-by-term integration is permissible. This gives (3).

LEMMA 5. *If $M(u)$ is continuous on the contour C surrounding the origin, then*

$$y(x) = \frac{1}{2\pi i}\int_C M(u)e^{ux}du$$

is of exp. val. $\leqq$ maximum value of $|u|$ on C.

We have $|M(u)| \leqq l$ (constant) on C. Now use the fact that the power series for e^{ux} is uniformly convergent on C.

We consider solutions $(y_1, y_2, \cdots)$ of (III) which are of exp. val. $\leqq \mu < R$. If we apply Lemmas 1–4 to (III) we obtain

$$\text{(a)} \qquad \sum_{n=0}^\infty \frac{1}{2\pi i}\int_C \frac{Y_n(u)}{u} A_{in}\left(\frac{1}{u}\right) e^{x/u}du = f_i(x),$$

and on integrating term-wise (formally, as yet) we arrive at the system of contour integral equations

$$\text{(III')} \qquad \frac{1}{2\pi i}\int_C \frac{e^{x/u}}{u}\left\{Y_1(u)A_{i1}\left(\frac{1}{u}\right) + \cdots + Y_n(u)A_{in}\left(\frac{1}{u}\right) + \cdots\right\} du = f_i(x) \qquad (i = 1, 2, \cdots),$$

where C is a contour, surrounding the origin and lying in $1/R < |u| < 1/\mu$.*

Now equations (III') represent the $f_i(x)$ as linear expressions in the $Y_n(x)$'s (under an integral). We may then hope to solve by interchanging the rôles of f, Y (and consequently using F, y). We look, then, for a solution of (III) in the form

---

* If we deal with the case of a finite system: (b) $\sum_{j=1}^n A_{ij}[y_j] = f_i(x)$, $i = 1, \cdots, n$, the step from (a) to (III') is no longer merely formal, and we have the following theorem of equivalence:

THEOREM. *If $(y_1, \cdots, y_n)$ of exp. val. $\leqq \mu < R$ is a solution of* (b), *then $(Y_1, \cdots, Y_n)$ is a solution of*

$$\text{(b')} \qquad \frac{1}{2\pi i}\int_C \left\{\sum_{j=1}^n Y_j(u)A_{ij}(1/u)\right\}\frac{e^{x/u}}{u}du = f_i(x) \qquad (i = 1, \cdots, n)$$

if the contour C, surrounding the origin, is chosen to lie in $1/R < |t| < 1/\mu$. Conversely, if $(Y_1, \cdots, Y_n)$, analytic in $|t| < 1/\mu$ (where $\mu < R$), is a solution of (b'), *then $(y_1, \cdots, y_n)$ is a solution of* (b) *and is of exp. val. $\leqq \mu$.*

$$y_i(x) = \frac{1}{2\pi i}\int_\Gamma \frac{e^{x/u}}{u}\Big\{F_1(u)\Lambda_{i1}\Big(\frac{1}{u}\Big) + \cdots + F_n(u)\Lambda_{in}\Big(\frac{1}{u}\Big) + \cdots\Big\}\,du \quad (i = 1,2,\cdots), \tag{4}$$

the contour Γ and the functions $\Lambda_{ij}(u)$ to be determined.

If we substitute the values (4) into (III) formally, we find the equations

$$f_i(x) = \frac{1}{2\pi i}\int_\Gamma \frac{e^{x/u}}{u}\Big\{F_1(u)\sum_{j=1}^{\infty}A_{ij}\Big(\frac{1}{u}\Big)\Lambda_{j1}\Big(\frac{1}{u}\Big) + \cdots + F_n(u)\sum_{j=1}^{\infty}A_{ij}\Big(\frac{1}{u}\Big)\Lambda_{jn}\Big(\frac{1}{u}\Big) + \cdots\Big\}\,du.$$

A sufficient condition (see Lemma 1) that these relations hold, $i=1, 2, \cdots$, is that the coefficient of $F_n(u)$ be identically unity or zero according as $n=i$ or $n\neq i$. This gives rise to the following infinite systems of linear equations in the unknowns $\Lambda_{jn}(u)$:

$$\sum_{j=1}^{\infty}A_{ij}(u)\Lambda_{jn}(u) = \delta_{in} = \begin{cases}0, i \neq n,\\ 1, i = n.\end{cases} \tag{5}$$

Here $i=1, 2, \cdots$, while n may be given a fixed value. As we vary n we obtain different systems of equations, but they all have the same determinant:

$$\Delta(u) = \begin{vmatrix} A_{11}\cdots A_{1n}\cdots \\ A_{21}\cdots A_{2n}\cdots \\ \cdot\quad\cdot \\ \cdot\quad\cdot \\ \cdot\quad\cdot \end{vmatrix}. \tag{6}$$

On solving systems (5) in the customary way (by determinants) we find that

$$\Lambda_{jn}(u) = \frac{\Delta_{jn}(u)}{\Delta(u)}, \tag{7}$$

where

$$\Delta_{jn}(u) = \begin{vmatrix} A_{11}\cdots A_{1,j-1} & 0 & A_{1,j+1}\cdots \\ \cdot\quad\cdot & \cdot & \cdot\quad\cdot \\ \cdot\quad\cdot & \cdot & \cdot\quad\cdot \\ & 0 & \\ A_{n1}\cdots A_{n,j-1} & 1 & A_{n,j+1}\cdots \\ \cdot\quad\cdot & 0 & \cdot\quad\cdot \\ \cdot\quad\cdot & \cdot & \cdot\quad\cdot \end{vmatrix}. \tag{8}$$

Returning to (4) we see that the brace in the integrand can be written as a determinant, so that we have

$$y_i(x) = \frac{1}{2\pi i}\int_\Gamma \frac{e^{x/u}}{u\Delta(1/u)}\Delta_i(u;F)du \quad (i = 1,2,\cdots,\infty), \tag{9}$$

where

$$\Delta_i(u;F) = \begin{vmatrix} A_{11}\left(\frac{1}{u}\right)\cdots A_{1,i-1}\left(\frac{1}{u}\right)F_1(u)\,A_{1,i+1}\left(\frac{1}{u}\right)\cdots \\ A_{21}\left(\frac{1}{u}\right)\cdots A_{2,i-1}\left(\frac{1}{u}\right)F_2(u)\,A_{2,i+1}\left(\frac{1}{u}\right)\cdots \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \end{vmatrix}. \tag{10}$$

We choose the contour Γ to surround the origin and to lie in $1/R < |u| < 1/\lambda$.

The above work has been formal. It can be made rigorous on the basis of our groups of assumptions (B) and (C). But we need only the final result (9). In fact, we have

THEOREM. *The functions $y_i(x)$ given by* (9) *satisfy system* (III) *and are of exp. val. $\leqq R' < R$.*

From conditions (B) and (C) it follows that $\Delta(1/u)$, $\Delta_i(u;F)$ converge uniformly* and absolutely on Γ. The functions $y_i(x)$ as given by (9) are then of exp. val. $\leqq R' = 1/r$ where r is the minimum distance of the origin from Γ (Lemma 5). Then, $R' < R$. Now

$$A_{ij}[y_j] = \frac{1}{2\pi i}\int_\Gamma \Delta_j(u;F)A_{ij}(1/u)\frac{e^{x/u}}{u\Delta(1/u)}du,$$

so that

$$(\alpha) \qquad \sum_{j=1}^{\infty} A_{ij}[y_j] = \frac{1}{2\pi i}\int_\Gamma \frac{e^{x/u}}{u\Delta(1/u)}\left\{\sum_{j=1}^{\infty} A_{ij}(1/u)\Delta_j(u;F)\right\}du,$$

provided the series in the brace is uniformly convergent on Γ. Let us establish this result. We can write

$$\Delta_i(u;F) = \sum_{j=1}^{\infty}\Delta_{ij}(1/u)F_j(u), \tag{11}$$

where Δ_{ij} is given by (8), and this series is absolutely and uniformly convergent on Γ; furthermore, the sum of the series of absolute values is bounded

* With respect to uniformity, here and later, see von Koch, loc. cit., pp. 238–244, and Riesz, loc. cit., pp. 39–40. The absolute convergence is a well known result: Riesz, loc. cit., p. 26.

uniformly* in i: $\sum_{j=1}^{\infty} |\Delta_{ij}(1/u)F_j(u)| \leqq K$, independent of i. From this follows the convergence, and uniform convergence on Γ, of the series $\sum_{j,k} |\Delta_{jk}(1/u)A_{ij}(1/u)F_k(u)|$. For

$$\sum_{j,k} |\Delta_{jk}A_{ij}F_k| = \sum_j |A_{ij}| \left\{ \sum_k |F_k\Delta_{jk}| \right\} \leqq K \sum_j |A_{ij}|,$$

and this last series converges. We may then sum the double series $\sum_{j,k}\Delta_{jk}(1/u)A_{ij}(1/u)F_k(u)$ in any manner, and one such summation yields

$$\sum_j A_{ij}\left\{ \sum_k F_k\Delta_{jk} \right\} = \sum_j A_{ij}(1/u)\Delta_j(u;F).$$

The uniformity of convergence is not destroyed, so that (α) is established.

But another mode of summing gives $\sum_k F_k\{\sum_j A_{ij}\Delta_{jk}\}$, and for the brace we have the well known relation†

$$\sum_{j=1}^{\infty} A_{ij}(1/u)\Delta_{jk}(1/u) = \begin{cases} 0, & i \neq k; \\ \Delta(1/u), & i = k. \end{cases}$$

Hence

$$\sum_j A_{ij}(1/u)\Delta_j(u;F) = F_i(u)\Delta(1/u),$$

so that we can rewrite (α) to give

$$(\beta) \qquad \sum_{j=1}^{\infty} A_{ij}[y_j] = \frac{1}{2\pi i}\int_\Gamma F_i(u)\frac{e^{x/u}}{u}du.$$

By Lemma 1 the right hand member is precisely $f_i(x)$. That is, $(y_1, \cdots, y_n, \cdots)$ is a solution of (III).

Let us return to the system of contour integral equations (III′). The set $(Y_1, \cdots, Y_n, \cdots)$ which corresponds to $(y_1, \cdots, y_n, \cdots)$ as defined by (9) is given by‡

$$(12) \qquad Y_i(x) = \frac{1}{2\pi i}\int_\Gamma \frac{\Delta_i(u;F)du}{(u-x)\Delta(1/u)} \qquad (i = 1,2, \cdots).$$

COROLLARY. *The functions $Y_i(x)$ defined by* (12) *satisfy system* (III′).

Equations (a) follow from system (III), since the $y_i(x)$ are of exp. val. $\leqq \mu < R$, where μ is properly chosen. If now we substitute (12) into (a) we obtain

---

* Riesz, loc. cit., p. 26.

† Riesz, loc. cit., p. 27.

‡ For y_i and Y_i are related in the same way that $e^{x/u}$ and $1/(1-x/u)$ are related.

$$\sum_{n=0}^{\infty}\left(\frac{1}{2\pi i}\right)^2 \int_C \int_\Gamma \frac{e^{z/u}A_{in}(1/u)}{u(t-u)\Delta(1/t)}\Delta_n(t;F)dudt = f_i(x).$$

But $\sum_{n=0}^{\infty} A_{in}(1/u)\Delta_n(t;F)$ converges uniformly on C, Γ, so that we can integrate term-wise. This gives (III′).

Remark. The assumptions (B) and (C) have been used only to establish the validity of the formal processes. These hypotheses may be replaced by others. In fact we have the*

THEOREM. *The functions $y_i(x)$ of* (9) *satisfy system* (III) *and are of exp. val. $\leqq R' < R$, when assumptions* (B), (C) *are replaced by*

(B′) (i) $A_{ij}(t)$ *analytic,* $|t| < R$;
(ii) $f_i(x)$ *is of exp. val.* $\lambda_i < \lambda < R$;
(iii) $\sum_{i=1}^{\infty} |F_i(u)|^2$ *converges uniformly in* $|u| \leqq 1/\lambda$; *and*
(i) *in the determinant* $\Delta(u)$, *the series*

(C′)
$$\sum_{i,j=1,\, i\neq j}^{\infty} |A_{ij}(u)|^2$$

and the product $\prod_{i=1}^{\infty} |A_{ii}(u)|$ *converge for every u in* $|u| < R$ (*it follows that* $\Delta(u)$ *is absolutely converent*†);
(ii) *the convergence of the above product and series is uniform in every closed region lying in* $|u| < R$;

(iii) $$\Delta(u) \not\equiv 0.$$

It is only necessary to go over the proof of the preceding theorem, observing that each step can be justified under the new hypotheses.‡

3. THE HOMOGENEOUS CASE

We consider now the homogeneous system

(IV) $$A_{i1}[y_1] + A_{i2}[y_2] + \cdots + A_{in}[y_n] + \cdots = 0 \qquad (i = 1,2,\cdots),$$

with assumptions (B), (C) or assumptions (B′), (C′), and look for a solution similar to the expressions (9). We set

---

* I owe to Professors Hille and Tamarkin the observation that the methods already employed will be valid under the conditions (B′), (C′).

† See H. von Koch, *Sur le convergence des déterminants infinis*, Rendiconti del Circolo Matematico di Palermo, vol. 28 (1909), pp. 255–266.

‡ For the theory of determinants which are absolutely convergent, see the paper of von Koch which has just been cited.

$$(13)\quad y_i(x) = \frac{-1}{2\pi i}\int_{\Gamma'} \frac{e^{x/u}}{u\Delta\left(\frac{1}{u}\right)} \begin{vmatrix} A_{11}\left(\frac{1}{u}\right)\cdots A_{1,i-1}\left(\frac{1}{u}\right)P_1\left(\frac{1}{u}\right), A_{1,i+1}\left(\frac{1}{u}\right)\cdots \\ A_{21}\left(\frac{1}{u}\right)\cdots A_{2,i-1}\left(\frac{1}{u}\right)P_2\left(\frac{1}{u}\right), A_{2,i+1}\left(\frac{1}{u}\right)\cdots \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \end{vmatrix} du,$$

where Γ' is a small contour surrounding $t=1/\xi$, with $\Delta(\xi)=0$ and $|\xi|<R$, and $P_1(u)$, $P_2(u)$, $\cdots$ are arbitrary polynomials* which vanish at $u=0$.

On making the transformation $t=1/u$ we obtain

$$(14)\quad y_i(x) = \frac{1}{2\pi i}\int_{\Gamma} \frac{e^{tz}}{t\Delta(t)} \begin{vmatrix} A_{11}(t)\cdots A_{1,i-1}(t)P_1(t)A_{1,i+1}(t)\cdots \\ A_{21}(t)\cdots A_{2,i-1}(t)P_2(t)A_{2,i+1}(t)\cdots \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots \end{vmatrix} dt,$$

where Γ is a small contour surrounding $t=\xi$ and containing no other zero of $\Delta(t)$. We find that†

$$A_{i1}[y_1] + \cdots + A_{in}[y_n] + \cdots = \frac{1}{2\pi i}\int_{\Gamma} P_i(t)\frac{e^{tz}}{t}dt,$$

and this is zero since the integrand is analytic in and on Γ. Consequently, $(y_1, y_2, \cdots)$ as given by (14) is a solution of system (IV); and to each zero $t=\xi$ of $\Delta(t)$ (in $|t|<R$) there corresponds one or more solutions of (IV).‡

---

* They are also subject to the condition that they make the determinant in which they appear uniformly convergent on Γ'. This will be the case if $|P_i(u)|\leqq N$ uniformly in i and u, where u ranges over any closed region in $|u|<R$ (case (B)) or if $\sum_{i=1}^{\infty}|P_i(u)|^3$ converges uniformly in every closed region in $|u|<R$ (case (B′)).

† Convergence questions are treated as in §2.

‡ In Sheffer, Annals Paper I, we show that for a finite system there are precisely $s(\xi)$ linearly independent solutions corresponding to the zero $t=\xi$ of $\Delta(t)$ (in $|t|<R$), where $s(\xi)$ is the order of the zero.

PRINCETON UNIVERSITY,
PRINCETON, N. J.

ON GENERAL TOPOLOGY AND THE RELATION OF THE PROPERTIES OF THE CLASS OF ALL CONTINUOUS FUNCTIONS TO THE PROPERTIES OF SPACE*

BY

E. W. CHITTENDEN

INTRODUCTION

Successive generalizations of the theory of sets of points based in turn on the concepts distance, limit of a sequence, neighborhood, and limiting point or point of accumulation led to the recognition of the concept point of accumulation as fundamental for the general theory of abstract sets.† The first writer formally to suggest founding the theory of abstract sets on this basis was F. Riesz‡ who proposed a set of postulates characteristic of a very general type of abstract set. The existence of a definition of point of accumulation for the subsets E of an aggregate P is logically equivalent to the existence of a single-valued function

$$E' = K(E)$$

whose argument E ranges over the class $\mathfrak{S}$ of all subsets E of the class P and whose values are likewise elements of $\mathfrak{S}$. Thus the theory of abstract sets is a phase of the study of *set-valued set-functions*, in fact of those functions which transform an aggregate $\mathfrak{S}$ of sets of elements of a class P into an aggregate $\mathfrak{S}'$ of the same character.

Since analysis situs or topology is primarily concerned with invariants under homeomorphic transformations of a space into itself the general theory of abstract sets provides a basis for the most general treatment of topological problems. For this reason F. Hausdorff§ called a very general

* Presented to the Society, December 28, 1923, April 15, 1927, and April 7, 1928; received by the editors August 18, 1928.

† M. Fréchet: (I) *Sur la notion de voisinage dans les ensembles abstraits*, Bulletin des Sciences Mathématiques, (2), vol. 42 (1918), pp. 1–19; (II) *Esquisse d'une théorie des ensembles abstraits*, Sir Asutosh Mookerjee's Commemoration Volumes, II, Calcutta, 1922, pp. 333–394; (III) *Sur les ensembles abstraits*, Annales Scientifiques de l'Ecole Normale Supérieure, vol. 38 (1921), p. 342; (IV) *Les Espaces Abstraits et leur Théorie considérée comme Introduction à l'Analyse Générale*, Paris, Gauthier-Villars, 1928.

‡ *Stetigkeitsbegriff und abstrakte Mengenlehre*, Atti del 4 Congresso Internationale dei Matematici, Roma, 1910, vol. 2, p. 18.

§ *Grundzüge der Mengenlehre*, Leipzig, Veit, 1914, p. 213.

type of abstract set a topological space. For the same reason Fréchet* has proposed that the term *topological space* refer to the most general form of abstract set, the case in which the points of accumulation are defined by an arbitrary set-function. This terminology will be used throughout the following discussion. A topological space of the Hausdorff type will be referred to as a Hausdorff space.

The first part of the present paper may be regarded as a discussion of the following problem: Determine the extent of the general theory of topological spaces.† It is also a study of various set-valued set-functions which may be defined in terms of a postulated arbitrary set-function $K(E)$. By the aid of suitable definitions many important theorems of the theory of sets of points are extended to the most general topological space.

Several of the set-functions defined by a given set-valued set-function $K(E)$ are of particular interest. The function $V(E)$ of §9 is associated with the concept neighborhood and defines a neighborhood space associated with the given topological space. The class of *proper nuclear* points‡ of a set E defines a set-function discussed in §11. The importance of this set-function was first recognized by C. Kuratowski and W. Sierpinski§ who used it in the solution of the following problem of Fréchet: *Determine the most general class* (L) *in which the theorem of Borel is true.*

It is of interest to observe that while the Hedrick property,‖ *the derived set of every set is closed,* may not be present in a topological space, it is present in associated spaces discussed in §§8 and 12. This fact permits the generalization to general topological spaces of many classical theorems.

The second part of the paper is devoted to a study of covering theorems in general topological space.¶ Of particular interest is the formulation in §20 of two sets of necessary and sufficient conditions that a set possess the "any to finite" form of the property of Borel.

---

* Congrès des Sociétés Savantes, Dijon, 1924, and Comptes Rendus, vol. 180 (1925), p. 419. See IV, p. 166.

† The results presented in this part of the paper were obtained prior to the appearance and independently of the article *La notion de dérivée comme base d'une théorie des ensembles abstraits*, by W. Sierpinski, Mathematische Annalen, vol. 97 (1926), pp. 321–337. An examination of the work of Sierpinski shows that his article duplicates the results here presented in part only.

‡ In a Hausdorff space these points coincide with the *complete limiting* points of P. Alexandroff and P. Urysohn, Mathematische Annalen, vol. 92 (1924), p. 258, and in general topological space with the *complete interior limiting points* of T. H. Hildebrandt, Bulletin of the American Mathematical Society, vol. 32 (1926), p. 469.

§ Fundamenta Mathematicae, vol. 2 (1921), pp. 172–8.

‖ E. R. Hedrick, these Transactions, vol. 12 (1911), pp. 285–294.

¶ T. H. Hildebrandt has given (loc. cit) an excellent résumé of the literature on covering theorems entitled *The Borel theorem and its generalizations.*

In the third part of this paper we consider the relationship between the properties of the class C of all continuous functions on a topological space and the properties of the space. It is usual in the theory of abstract sets for writers to impose conditions on the space and deduce the properties of the class C which result. It is evident, however, that many properties of the class C cannot be present in spaces of unrestricted generality. For example, if the space is enumerable and connected every continuous function is constant. Thus the assumption of the existence of a non-constant function restricts the character of the space. Fréchet has proposed, in correspondence with P. Urysohn and myself, the following problem: Characterize the most general space in which there exists a non-constant continuous function.

The solution of this problem presented in §21 is a generalization of a result of P. Urysohn* stated for Hausdorff spaces.

A. D. Pitcher† has shown for the space of a symmetric écart‡ that if the class C of all continuous functions is *uniformly proper*§ the space is compact, connected, and metric. In this paper we consider a group of fundamental properties possessed by the class C of all continuous functions on a metric space, and obtain a number of topological conditions which are necessary for the presence of combinations of these properties in the case of a topological space. These results extend and supplement those obtained by Pitcher.

It is important to note that the definition of continuous function on a topological space given by Fréchet‖ is in fact a neighborhood definition, and that in consequence the theory of continuous functions on the classes (V) is as general as that obtained in the case of the most general topological space.

I. Derived set-functions

1. Set-functions. Throughout the following discussion the term *space* or *topological space* will be used to denote a system (P, K) composed of an

---

* Mathematische Annalen, vol. 94 (1925), p. 290.

† These Transactions, vol. 19 (1918), pp. 73–76.

‡ A symmetric non-negative function of pairs of points p, q, which is defined for all pairs of points of the class P, and vanishes if and only if $p=q$. Cf. Fréchet, these Transactions, vol. 19 (1918), p. 53–65.

§ A class of continuous functions is uniformly proper if every function of the class is bounded, uniformly continuous, assumes every value between its bounds, and has the following further properties: if p and q are distinct points there is a function $f(p)$ of the class such that $f(p) \neq f(q)$; if A and B are any two sets such that the écart of pairs of points one from each set has a positive minimum, there is a continuous function $f(p)$ such that $f(p)=0$, if p is in A, $=1$, if p is in B, and $0 \leqq f(p) \leqq 1$, for all other points.

‖ Loc. cit. III, p. 363.

abstract aggregate or set P and a relation $E'KE$ between subsets E, E' of P.

The set E' is assumed to be unique and determined for each subset E of the fundamental set P. Thus the relation $E'=K(E)$ defines a single-valued set-valued set-function, whose range is the class of all subsets of P, and whose values are also subsets of P. Different set-functions $K(E)$, $H(E)$, relative to the same set P determine different spaces.

2. **Properties of sets of points.** The elements of the class P in a space (P, K) will be called points. The elements, if any, of the set $E'=K(E)$ will be called *points of accumulation* of the set E with reference to the function $K(E)$, relation K, or space (P, K). The terms of the theory of sets of points retain their usual significance in relation to the points of accumulation defined by a set-function.

In case several set-functions are to be considered simultaneously we indicate the function of reference by writing K-point, K-closed, K-interior, etc.

3. **Derived set-functions.** In this and following paragraphs we consider a number of set-functions which may be defined in terms of a given set function $K(E)$, together with the corresponding spaces and types of points of accumulation.

The complement $P-E$ of a set E may be represented by the set function $C(E)$. In this notation the definition of interiority may be stated as follows. *A point p is K-interior to a set E if it is not a K-point of any subset of $C(E)$.** The set $I(E)$ of all the K-interior points of a set E is of great importance.

The function $I(E)$ is easily seen to be monotonic increasing. That is, if E is a subset of a set G, $I(E) \leqq I(G)$.

Open sets are defined by the equation $E=I(E)$, that is, they consist entirely of interior points. It is quite easy to show that the sum of a finite or infinite number of open sets is open. The corresponding proposition regarding the product of closed sets cannot be extended to general space.

The frontier of a set E, $F(E)$, is given by the formula

$$F(E) = E \cdot K[C(E)] + C(E) \cdot K(E).$$

The isolated points of a set E define the function $S(E)$.

4. **Iteration of set-functions.** By finite iteration we obtain from a given set-function $K(E)$ the derived sets and derived functions of finite order

$$E^{(\alpha)} = K^{(\alpha)}(E).$$

Then for ordinals of the second kind we define

$$E^{(\alpha)} = K^{(\alpha)}(E)$$

* M. Fréchet, loc. cit. II.

as the outer limit of $E^{(\beta)}$ where β varies over the set of all ordinals preceding α. Finally we set

$$E^{(\alpha+1)} = K^{(\alpha+1)}(E) = K[K^{(\alpha)}(E)].$$

Since the function K is general we have defined the iterates of finite and transfinite order for all set-valued set-functions.

A set-function may be cyclic under iteration. For example, the function $C(E) = P - E$ is cyclic of period two.

5. **Composition of set-functions.** From two set-functions G, H we may derive others by various forms of composition. For example, the sum $G+H$ and the product GH are defined by the equations

$$\overline{G + H}(E) = G(E) + H(E), \quad GH(E) = G(E) \cdot H(E).$$

We may also consider functions of functions, as $G[H(E)]$.

6. **Fundamental properties of set-functions.** For convenience of reference we present a list of the properties of set-functions which are fundamental for the present study.*

I. (Monotonic.) *If a set A is a subset of a set B, then $K(A)$ is a subset of $K(B)$.*

II. *If $E = A + B$, then $K(E) \leqq K(A) + K(B)$.*

III. *For every set E, $K''(E) = (K'E)$.*†

IV. *A set of one element has no K-point.*

V. *Each point of a set E is uniquely determined by the family of subsets of E of which it is a K-point.*

If a set-function has both of the properties I and II it is *additive*, that is

$$K(A + B) = K(A) + K(B).$$

It follows readily from the third property (III) that all the iterates of the function $K(E)$ have the property, and furthermore, that for any two ordinal numbers α, β $(\alpha < \beta)$,

$$K^{(\alpha)}(E) \leqq K^{(\beta)}(E).$$

Properties I and IV together imply that the null set has no K-point.

7. **Monotonic set-functions.** From every set-function $K(E)$ we derive a monotonic set-function $L(E)$ of considerable interest. The L-points of a

---

* Of these properties, I, II, IV, V are equivalent to the four properties of F. Riesz, loc. cit. The third corresponds to the Hedrick property (loc. cit.).

† A space with this property is said to be *accessible*. Cf. Fréchet, loc. cit. I.

set E are defined to be those points of the space (P, K) which are K-points of some subset of E. Obviously

$$K(E) \leqq L(E).$$

It is also evident that the function $L(E)$ is monotonic. *If the function $K(E)$ is monotonic, $K(E) = L(E)$ for every set E.*

The terms K-interior and L-interior are equivalent. Hence every open set in the space (P, K) is open in the derived space (P, L), and vice versa.

The property L-closed is stronger than the property K-closed. The complement of an open set in the space (P, K) is L-closed, and the complement of an L-closed set is open in both spaces. Because of these facts we shall in considering general topological spaces refer to the sets which are L-closed as *completely* closed. The notation F is reserved for the completely closed sets, preserving the usual formulas,

$$C(O) = F, \quad C(F) = O.$$

Since the sum of a family of open sets is an open set it follows at once that *the product of a finite or infinite family of completely closed sets in a general topological space is completely closed.*

8. **A derived set-function with the Hedrick property.** Let $M^0(E)$ denote the set common to all completely closed sets containing the set E as subset. This set exists (since the set P is completely closed) and is completely closed. The function $M^0(E)$ has the fundamental properties I and III. It will now be shown that this function may be defined in terms of the function $L(E)$ and the process of iteration.

Let a set-function $M(E)$ be defined by the equation

$$M(E) = E + L(E).$$

By iteration we obtain

$$M''(E) = M(E) + LM(E).$$

The series of functions $M^{(\alpha)}(E)$ derived from $M(E)$ by iteration has the property that if α precedes β then $M^{(\alpha)}(E)$ is a subset of $M^{(\beta)}(E)$. If we accept the theory of well ordered series in general, there must exist an ordinal for which

$$M^{(\alpha+1)}(E) = M^{(\alpha)}(E).$$

The set $M^0(E) = M^{(\alpha)}(E)$ so defined is the least completely closed set containing E. Evidently the iterates of the function $M^0(E)$ are identical to M^0.

But $M^0(E)$ will in general contain points of E which are not points of accumulation in the usual sense. However we may say that a point p is a

point of accumulation of a set E of *degree* α provided α is the least ordinal for which the point p is an L-point of $M^{(\alpha)}(E)$. The set of all points which are points of accumulation of E of degree α for some value of α will be denoted by $L^0(E)$. This set is also completely closed. We have thus obtained the closure of derived sets in another manner.

THEOREM. *Every topological space* (P, K) *determines an accessible space* (P, L^0). *The spaces coincide if the function* $K(E)$ *has properties* I *and* II. *For all values of the ordinal* α

$$K^{(\alpha)}(E) \leqq L^0(E).$$

If a set E *is closed or open in the space* (P, L^0) *it is completely closed, or open, in any of the spaces* (P, K), (P, L), $(P, K^{(\alpha)})$, $\alpha = 1, 2, 3, \cdots$.

The significance of this theorem in the case of certain singular set-functions is illustrated by the example $K(E) = C(E)$. Then $L^0(E) = P$.

9. **Neighborhood spaces.** Any set V of points of a space (P, K) which has a point p for an interior point will be called a *neighborhood* of p. Then the set function $V(E)$ may be defined by the set of all points p such that every neighborhood of p contains a point of E distinct from p. The function $V(E)$ is monotonic increasing and has the further property

IIa. *If a point* p *is a* V*-point or neighborhood point of a set* E, *it is a neighborhood point of the set* $E - p$.

The space (P, V) so defined is a class (V) of Fréchet.* It has been shown by Fréchet that properties I and IIa are together a necessary and sufficient condition that a function $K(E)$ be identical to the derived function $V(E)$.

Every neighborhood point of a set E is an L-point. If a point p is a K-point of the set $E - p$ it is a neighborhood point of E. If a point p is a K-point of a set E, but is not a K-point of the set $E - p$ or any subset of that set, it is not a V-point of E. The K-points which are not neighborhood points are virtually isolated, and will be called *singular* points of accumulation.

The neighborhoods derived from the function $V(E)$ are identical to those derived from the function $K(E)$.

If the null set has no L-point, then each point has one neighborhood at least (which may be the aggregate P) and every neighborhood of a point p contains p. This is property A of the Hausdorff postulates for a neighborhood space.†

If in the definition of the function $V(E)$ we replace the family of all

* Loc. cit. I.

† Loc. cit., p. 213.

neighborhoods of a point p by the family of all open sets O which contain p we obtain a function $O(E)$. Clearly $V(E)=O(E)$. The function $O(E)$ has properties I, IIa, and III. A necessary and sufficient condition that the functions $V(E)$ and $O(E)$ be identical is that for each point p there exist a neighborhood V such that $M^0(P-V)$ does not contain the point p. It is quite easy to show that this is the case if the function $K(E)$ has the properties I, II, and III.

10. **Nuclear points.** Hausdorff* has introduced the class E_μ of all points p of order μ at least relative to E; that is, the class of all points such that every neighborhood contains a subset of E of power μ. It is evident that if $\mu<\nu$ then $E_\mu \geqq E_\nu$. This definition defines a series of functions denoted by $V_\mu(E)$ $(\mu=1, 2, 3, \cdots)$. A point of order μ will be called a μ-point. If it is necessary for the sake of clearness, we may write K_μ-point, L_μ-point etc.

If a set E contains the set $V_\mu(E)$ it is μ-closed.

Since each neighborhood of a point p contains p we have at once

$$E + K_1(E) = K_1(E) = L_1(E) = V_1(E).$$

If $\mu=1$, every μ-point is an L-point. The case $\mu=0$ is trivial.

The μ-points of a set E of order $\mu=|E|$, the cardinal number of E, are called *nuclear*† points since they are V-points of the nucleus of E, the set of all points of E which are of order $|E|$ relative to E. Every nuclear point of a set of more than one element is an L-point.

If every infinite subset of a set E has a nuclear point in E, the set E is said to be self-nuclear. *Every self-nuclear set is self-compact relative to the function* $L(E)$.

11. **Proper points of accumulation.** A point p is a *proper* point of accumulation of a set E if for every neighborhood V of p there is a point q of E interior to V.‡ Every proper point of accumulation is a V-point. A K-point which is not a proper point of accumulation is said to be *improper*.

If the system of neighborhoods is equivalent to a family of open sets every point of accumulation is proper.

Among the proper points of accumulation are the points of *power* μ§ which are related in an interesting and remarkable manner to the covering theorems.

---

* Loc. cit., p. 219 ff.

† E. W. Chittenden, Bulletin of the American Mathematical Society, vol. 30 (1924), p. 514. They correspond to the complete points of accumulation of Alexandroff and Urysohn in the case of a Hausdorff space (loc. cit.).

‡ Cf. T. H. Hildebrandt, loc. cit.

§ The terms "order" and "power" are due to Hildebrandt, loc. cit.

A point p is of *power* μ at least relative to a set E if every neighborhood of p contains a subset of E of power μ in its *interior*.

If $\mu > 0$, the points of power μ are also of order μ, and may therefore be called proper μ-points. The class of all proper μ-points of a set E is denoted by $H_\mu(E)$. A proper V-point is therefore an H_1-point.

The points of power $|E|$ relative to E are called *proper nuclear* points.* If a set E contains a proper nuclear point of every infinite subset of E, E is said to be *properly self-nuclear*.

If every infinite subset of a set E determines a proper K-point in E, E is *properly self-compact*.† In a space (P, L) every properly self-nuclear set E is properly self-compact.

A set E is *properly closed* if it contains all its proper points of accumulation. This closure is understood to be relative to the type of point of accumulation under consideration.

THEOREM. *The class $H_2(E)$ is properly closed in the space (P, L).*

A point p is a proper interior point of a set E in case every set A which has p for a V-point contains an ordinary interior point of E which is distinct from p. Let $D(E)$ denote the set of all proper interior points of E. The points of $I(E)-D(E)$, if any, are K-points of some subset of $P-I(E)$, but not of $P-E$. They may however be points of accumulation of the set $C(E)$ of some degree greater than the first.

We note the following evident propositions:

Every point of an open set is a proper interior point.

The set $D(E)$ is interior to $I(E)$.

In the space (P, H_1) the set $I(E)$ is open.

12. **Other set-functions.** In case μ is equal to the cardinal number of the positive integers the function $H_\mu(E)$ has the first four of the fundamental properties listed in §6. The corresponding space is therefore a class (H) of Fréchet.‡ For this reason we denote the set $H_\mu(E)$, when $\mu = \aleph_0$, by $H(E)$. It follows that all the theorems which have been established for classes (H) may be extended when suitably modified to general topological space.

Another set-function of this kind determines an accessible space. A point p will be called a J-point of a set E if the interior of each neighborhood of p contains at least two points of E.

---

* In the article cited above I called this type of point hypernuclear.

† On page 195 (loc. cit. IV) Fréchet introduces the concept compact in "sens très strict" which is equivalent for classes (V) to properly compact. Many of the theorems stated by Fréchet are extended in following sections to general topological space.

‡ *Esquisse*, p. 354 and p. 366.

It follows at once that $J(E)=V(E)$, that is, that every J-point is a V-point. But the space (P, J) does not in general satisfy the condition IIa, and so need not be a neighborhood space. However, the function $J(E)$ has the fundamental properties I, III, and IV. But a function $K(E)$ with these properties is not necessarily equivalent to the corresponding derived function $J(E)$. The following proposition which is an immediate consequence of the definition of the function $J(E)$ is of interest.

The interior $I(E)$ of a set E in a general topological space is an open set in the derived space (P, J).

II. Covering theorems in general topological space

13. **Properties of coverings.** A family $\mathfrak{F}$ of sets G in a space (P, K) is a covering of a set E if every point p of E is an element of some set G of $\mathfrak{F}$; a proper covering in case each point p is *interior* to some set G. More generally a covering $\mathfrak{F}$ is of type T if each point p is in a relation T to some set G of $\mathfrak{F}$. From the axiom of choice we may conclude that every covering $\mathfrak{F}$ contains a covering $\mathfrak{F}_1$ of the same type and of cardinal number $|E|$ at most.

A covering $\mathfrak{F}$ of a set E is *reducible* if it contains a covering of the same type and of lower cardinal number, otherwise it is irreducible.

A covering $\mathfrak{F}$ of type T and power μ is *normal* if it can be arranged in a normally ordered series

$$G_1, G_2, G_3, \cdots, G_\alpha, \cdots \qquad (\alpha < \Omega_\mu),$$

where Ω_μ is the first ordinal of power μ, and there is for each of the ordinals α a point q_α of E in the relation T to G_α, which is not in that relation to any G_β, where $\beta<\alpha$. The set $Q=[q_\alpha|\alpha<\Omega_\mu]$ will be said to be *associated* with the normal covering $\mathfrak{F}$. Every normal covering is irreducible. It is evident that *the set of points of Q in the relation T to G_α is of power less than μ for every $\alpha<\Omega_\mu$.*

Theorem. *Every covering $\mathfrak{F}$ of a set E contains a normal covering of E of the same type.*

Let $\mathfrak{F}_1$ be any irreducible subfamily of $\mathfrak{F}$ which covers a given set E and let

$$G_1, G_2, G_3, \cdots, G_\alpha, \cdots \qquad (\alpha < \Omega_\mu) \tag{1}$$

represent any arrangement of the sets G of the family $\mathfrak{F}_1$ in a well ordered series of type Ω_α. Let α_1 be the first of the indices α for which there is a point p_1 of E in the relation T to G_α. In general G_{α_β} contains a point p_β of E not in the relation T to G_{α_γ} if $\gamma<\beta$. The index β must range over an ordinal series

coterminal with that of α, therefore ordinally similar to that series, since the family $\mathfrak{F}_1$ is irreducible. The family

$$\mathfrak{F}_2 = [G_{\alpha_\beta} | \beta < \Omega_\mu]$$

is easily seen to be a normal covering of E of type T and power μ.

14. **The theorem of Borel in a Hausdorff space.** P. Alexandroff and P. Urysohn* have formulated for Hausdorff spaces the following elegant generalizations of the classical theorem of Borel.

THEOREM I_0. *In a Hausdorff space the following properties of a set E are equivalent.*

A_0. *Every enumerable subset of E has a nuclear point.*

B_0. *The product of every enumerable decreasing sequence of non-null closed subsets of E contains a point of E.*

C_0. *Every enumerably infinite family of open sets which covers E is reducible.*

A necessary and sufficient condition that a set E possess these three properties is that it be self-compact.

THEOREM II. *In a Hausdorff space the following properties of a set E are equivalent.*

A. *Every infinite subset of E has a nuclear point in E.*

B. *The product of every well ordered decreasing sequence of non-null closed subsets of E contains a point of E.*

C. *Every infinite family of open sets which covers E is reducible.* †

A set E with one of the properties A, B, C has them all and is called *bicompact* by Alexandroff and Urysohn.

Alexandroff and Urysohn have shown that the properties A_0, B_0, C_0 are the first of a series of sets of equivalent properties A_μ, B_μ, C_μ, obtained by replacing the word "enumerable" in the statement of Theorem I_0 by the phrase "of power μ at most." If μ is the power of the set E we obtain property A, etc.

This generalization may be regarded as involving a raising of the upper limit of the powers of the sets entering the conditions A_0, B_0, C_0. A different generalization is obtained by raising the lower limit. Thus property C_0 might be changed to read as follows.

---

* Mathematische Annalen, vol. 92 (1924), p. 259. For a detailed discussion of this and related theorems see the article by Hildebrandt already cited. See also Tychonoff and Vedenissoff, Bulletin des Sciences Mathématiques, vol. 50 (1926), pp. 15–37.

† Because a finite covering of E results after a finite number of steps this statement of property C is equivalent to that usually given.

C_μ. *Every family of open sets of power μ which covers E is reducible.*

These considerations may be made still more general by limiting the range of powers above and below. For example if $\lambda \leqq \mu \leqq \nu$ are infinite cardinals we may consider the property

$C(\lambda \leqq \mu \leqq \nu)$. *Every family of open sets of power μ, $\lambda \leqq \mu \leqq \nu$, which covers E is reducible; therefore contains a subfamily which covers E and is of power less than λ.*

The property of Lindelöf, *Every infinite family of open sets which covers E contains an enumerable subfamily which also covers E,** is a special case of the property $C(\lambda < \mu \leqq \nu)$.

We proceed to extend the foregoing considerations to topological spaces in general. In this investigation properties analogous to A, B, C are fundamental. We first develop a series of propositions relative to coverings of an abstract type T.

15. **A theorem on reducibility of coverings.** Let μ be any cardinal number and let $\mathfrak{F} = [V_\alpha \mid \alpha < \Omega_\mu]$ be any infinite normal covering of E of type T and power μ. Denote the corresponding associated subset of E by $Q = [q_\alpha]$. Then $|Q| = \mu$, and the sequence

$$\mathfrak{S} = [G_\alpha], \quad G_\alpha = \sum_{\beta \geqq \alpha} q_\beta,$$

is of power μ. Furthermore there is, by the definition of Q, no set V of $\mathfrak{F}$ in the relation T to q_β if $\beta > \alpha$. Consequently no set V_α can be in the relation T to a point of every set G_α of $\mathfrak{S}$.

The result just obtained is formulated in the following lemma.

LEMMA. *If T is a relation between a set V and an element of V, μ is a cardinal number, and $\mathfrak{F} = [V]$ is an irreducible covering of a set E of type T and power μ, then*

(A) *there is a subset of Q of E of power μ such that no set V of the family $\mathfrak{F}$ is in the relation T to μ elements of E.*

This implies further that

(B) *there is a decreasing sequence $\mathfrak{S} = [G]$ of subsets of E such that $|S| = \mu$ and no set V of $\mathfrak{F}$ is in the relation T to an element of every set G of $\mathfrak{S}$.*

From this lemma we readily conclude the following fundamental theorem:

* Called the property of perfect separability by Fréchet. See Tychonoff and Vedenissoff, loc. cit., for a discussion of spaces (P, V) with this property.

THEOREM. *If μ is any cardinal number, T any relation between a set V and one of its elements, $\mathfrak{F} = [V]$ any covering of a set E of type T, the property*

(A) *there exists for every subset of Q of E of power μ a set V of $\mathfrak{F}$ in the relation T to μ points of Q,*

is implied by the property

(B) *there is for every decreasing sequence $\mathfrak{S} = [G]$ of subsets of E, such that $|\mathfrak{S}| = \mu$, a set of V of $\mathfrak{F}$ in the relation T to an element of every set G of $\mathfrak{S}$, and both properties imply*

(C) *the covering $\mathfrak{F}$ is either reducible or of power less than μ.*

16. **Decreasing sequences of sets.** The rôle of decreasing sequences of sets of points in the analysis of covering theorems was first exhibited for the classes (L) of Fréchet by R. L. Moore.* The symbol $\mathfrak{S} = [G]$ will be adopted as a fixed notation for a decreasing sequence of sets G of elements. The sequence $\mathfrak{S}$ is subject to the condition that given any two sets of $\mathfrak{S}$, one is a proper subset of the other. Suppose the sets G of $\mathfrak{S}$ are arranged in normal order. Evidently there will exist well ordered sequences in $\mathfrak{S}$ of the form

$$\mathfrak{S}_0 = [G_\alpha \mid \alpha < \Omega_\mu]$$

with the following property: every set G of $\mathfrak{S}$ contains all the elements of some set G_α of $\mathfrak{S}_0$. The sequences $\mathfrak{S}$ and $\mathfrak{S}_0$ are said to be *equivalent.*

Among all the well ordered sequences $\mathfrak{S}_0$ equivalent to a given sequence $\mathfrak{S}$ is a class of sequences of minimal cardinal number μ. This cardinal is regular, as the contrary hypothesis permits the existence of an equivalent sequence of lower cardinal number. The cardinal μ so determined will be called the regular cardinal belonging to the sequence $\mathfrak{S}$.

Let T be any relation between a set V and one of its elements. A sequence $\mathfrak{S} = [G]$ is *closed* in a set E with respect to a relation T if *there is a subset V of E which is in the relation T to an element of every set G of $\mathfrak{S}$.*

It is evident that if a sequence $\mathfrak{S}$ is closed the corresponding minimal sequences $\mathfrak{S}_0$ are also closed and vice versa.

If p is point of E common to every set G of a sequence $\mathfrak{S}$ and there is a set V in the relation T to p, then $\mathfrak{S}$ is closed in E by definition. For this reason we restrict the term *open sequence* to the case in which no element of E is common to all the sets G.

Let $\mathfrak{S} = [G]$ be an open sequence in a set E. There is a set $Q = [q_\alpha]$ of regular power μ such that for every $\alpha < \Omega_\mu$ there is a set G which does not

* Proceedings of the National Academy of Sciences, vol. 5 (1919), pp. 206–210.

contain q_α. Furthermore there is no set V in the relation T to μ points of the set Q. Otherwise V would be in the relation T to an element of every set G of the sequence $\mathfrak{S}$. This proves the following theorem:

THEOREM. *If for every set Q of elements of a set E of regular cardinal number μ there is a subset of V of E in the relation T to μ elements of Q then every decreasing sequence $\mathfrak{S}$ of subsets of E whose regular cardinal is μ is closed in E.*

This theorem has the following corollary:

THEOREM. *If for every infinite set Q of elements of a set E there is a subset V of E in the relation T to $|Q|$ elements of Q, then*

(B) *every infinite decreasing sequence of subsets of E is closed in E with respect to the relation T.*

17. **Decreasing sequences derived from normal coverings.** Let $\mathfrak{F} = [V_\alpha | \alpha < \Omega_\mu]$ be an irreducible normal covering of a set E of type T and power μ. Let Ω_ν be the ordinal series of least cardinal number ν coterminal with Ω_μ. The regular cardinal ν is said to be the regular cardinal of the series Ω_μ, or of the cardinal μ. If μ' is a cardinal whose ordinal series $\Omega_{\mu'}$ is coterminal with Ω_μ it is easy to see that ν is the corresponding regular cardinal.

There is a subset $Q = [q_\beta | \beta < \Omega_\nu]$ of E such that q_β is in the relation T to V_{α_β} and not in that relation to V_α for any $\alpha < \alpha_\beta$. The indices α_β form a series coterminal with the series $\alpha < \Omega_\mu$. The cardinal number of Q is ν. Consider the decreasing sequence of subsets of E,

$$\mathfrak{S} = [G_\beta | \beta < \Omega_\nu], \quad G_\beta = \sum_{\gamma > \beta} q_\gamma.$$

Since no set V_α can contain an element of every set G_β we have the following theorem:

THEOREM. *If $\mathfrak{F}$ is an irreducible covering of a set E of type T and if ν is the corresponding regular cardinal there exists a decreasing sequence $\mathfrak{S} = [G]$ of subsets of E such that $|\mathfrak{S}| = \nu$, and no set V of $\mathfrak{F}$ is in the relation T to an element of every set G.*

The following two propositions are immediate consequences.

THEOREM. *If $\mathfrak{F}$ is an irreducible covering of a set E of type T, ν is a regular cardinal number, and if for every decreasing sequence $\mathfrak{S} = [G]$ of subsets of E such that $|\mathfrak{S}| = \nu$ there is a set V in $\mathfrak{F}$ which is in the relation T to an element of each set G of $\mathfrak{S}$, then ν is not the regular cardinal determined by the cardinal number of $\mathfrak{F}$.*

THEOREM. *If $\mathfrak{F}$ is any infinite covering of a set E of type T, and if for every infinite decreasing sequence $\mathfrak{S}=[G]$ of subsets of E there is a set V in $\mathfrak{F}$ which is in the relation T to an element of each set G of the sequence $\mathfrak{S}$, then $\mathfrak{F}$ is reducible to a finite covering.*

No set V of the family $\mathfrak{F}$ can contain ν elements of the set Q. For by the definition of q_β, V_α is not in the relation to q_{α_β} if $\alpha<\alpha_\beta$. Hence the elements of Q in the relation T to V_{α_β} form a subset of $Q-G_\beta$ which is of power less than ν because ν is a regular cardinal. This result is formulated in the following theorem.

THEOREM. *If $\mathfrak{F}=[V]$ is an irreducible covering of a set E of type T, ν is any regular cardinal number, and if for every subset Q of E of power ν there is a set V of $\mathfrak{F}$ in the relation T to ν elements of Q, then the cardinal ν is not the regular cardinal determined by the cardinal number of the family $\mathfrak{F}$.*

From this theorem we obtain as a corollary the following fundamental proposition:

THEOREM. *If $\mathfrak{F}$ is any covering of a set E of type T and if for every infinite subset Q of E there is a set V in the family $\mathfrak{F}$ which is in the relation T to $|Q|$ elements of Q then $\mathfrak{F}$ may be reduced to a finite covering.*

18. **On the power of an irreducible covering.** The following conditions are sufficient to ensure that the irreducible coverings of a given type T shall be of power less than a given infinite cardinal.

THEOREM. *Let $\mathfrak{F}=[V]$ be any irreducible covering of a set E of type T, and let λ be any infinite cardinal. Then if*

(A) *for every subset Q of E of power λ at least there is a set V of $\mathfrak{F}$ in the relation T to $|Q|$ points of Q,*

or if

(B) *for every decreasing sequence $\mathfrak{S}=[G]$ of subsets of E for which $|\mathfrak{S}|\geqq\lambda$, there is a set V of $\mathfrak{F}$ in the relation T to an element of each set G of $\mathfrak{S}$,*

it follows that the power of $\mathfrak{F}$ is less than λ. Furthermore, property B *implies property* A.

The conclusion will read $|\mathfrak{F}|\leqq\lambda$ if we replace the conditions $|Q|\geqq\lambda$, $|\mathfrak{S}|\geqq\lambda$, by $|Q|>\lambda$, $|\mathfrak{S}|>\lambda$.

The family $\mathfrak{F}$ may be assumed to be in the irreducible normal form

$$\mathfrak{F}=[V_\alpha],$$

where the index α ranges over the first well ordered series of cardinal number $|\mathfrak{F}|$. Let $Q=[q_\alpha]$ be the subset of E associated with $\mathfrak{F}$. Then $|Q|=|\mathfrak{F}|$.

If $|Q| \geqq \lambda$ there is by (A) a set V_α of $\mathfrak{F}$ which is in the relation T to $|Q|$ points of Q. As this is contrary to the definition of Q as an associated set for the irreducible family $\mathfrak{F}$ it follows that the power of Q, and therefore of $\mathfrak{F}$, is less than λ.

To establish the second part of the theorem let

$$G_\alpha = \sum_{\beta > \alpha} q_\beta .$$

The sequence $\mathfrak{S} = [G_\alpha]$ cannot be of cardinal number λ or greater since there must exist in that case a set V_{α_0} of $\mathfrak{F}$ which is in the relation T to an element of every G_α, therefore to an element of G_β for $\beta < \alpha_0$. But q_β is by definition not in the relation T to any set V_α for $\alpha < \beta$.

That property (B) implies (A) can be shown quite easily by assuming that the elements of a set Q of power λ at least are arranged in a well ordered series of minimal ordinal type. From this series we obtain a sequence $\mathfrak{S}$ as in previous arguments. The set V of $\mathfrak{F}$ which contains a point of every set G of $\mathfrak{S}$ must contain $|Q|$ points of Q.

19. **Necessary and sufficient conditions that every infinite covering of type T be reducible.** In the preceding paragraphs we obtained sufficient conditions for the reducibility of coverings of power λ at least. The conditions become necessary if $\lambda = \aleph_0$.

THEOREM. *If E is any aggregate and $\mathfrak{F} = [V]$ is any normal covering of E of type T, the following properties are equivalent:*

(A) *Every infinite subset A of E contains a subset B of power $|B| = |A|$ such that every element of B is in the relation T to a single set V of $\mathfrak{F}$.*

(B) *For every infinite decreasing sequence $\mathfrak{S} = [G]$ of subsets of E there is a set V of $\mathfrak{F}$ in the relation T to an element of each set G of $\mathfrak{S}$.*

(C) *The covering $\mathfrak{F}$ is finite.*

From the theorem of §15 we have at once B→A→C. It is therefore sufficient to prove that C→B.

Let $\mathfrak{S} = [G]$ be an infinite decreasing sequence of subsets of E and let $\mathfrak{F} = [V_1, V_2, \cdots, V_n]$ be a finite covering of E of type T. Let G_i denote a set of $\mathfrak{S}$ which has no element in the relation T to a set V_i of $\mathfrak{F}$ ($i = 1, 2, \cdots, n$). Since the sets G_i form a finite monotone sequence they have a common element p. Since this element cannot be in the relation T to any set V_i of $\mathfrak{F}$ we have contradicted the hypothesis that $\mathfrak{F}$ covers E of type T.

20. **The theorem of Borel in general topological space.** If we consider the special case in which the relation T means "interior to" in the sense of the foregoing discussion, it is at once clear that the family $\mathfrak{F}$ of all possible

neighborhoods of the points of E is a covering of E of type T. In this case the theorem of the preceding section assumes the following form:

THEOREM. *If E is any set of points in a topological space and $\mathfrak{F}$ is any infinite proper covering of E the following properties are equivalent:*

(A) *Every infinite subset A of E contains a subset B of power $|B| = |A|$ such that B is interior to some set V of $\mathfrak{F}$.*

(B) *For every infinite decreasing sequence $\mathfrak{S} = [G]$ of subsets of E there is a set V of $\mathfrak{F}$ which has an element of every set G of $\mathfrak{S}$ in its interior.*

(C) *The covering $\mathfrak{F}$ is reducible.*

From this proposition we easily obtain the following generalization of the theorem of Borel.

THEOREM. *In a general topological space the following three properties of a set E are equivalent:*

(A) Sierpinski-Kuratowski. *Properly self-compact. Every infinite subset of E has a proper nuclear point in E.*

(B) R. L. Moore. *Perfectly properly self-compact. Every infinite decreasing sequence of subsets of E is closed in E.*

(C) Borel. *Bicompact. Every infinite proper covering of E is reducible.*

With this proposition the second theorem of Alexandroff and Urysohn cited in §14 above receives a final generalization.

21. **Perfectly compact sets in a topological space.** We have shown that the generalization of the theorem of Borel to topological space leads to the introduction of the property *perfectly properly self-compact* instead of the property perfectly self-compact. The two properties coincide in any V-space in which derived sets are closed, in particular in a Hausdorff space. The following results are presented for comparison with those of the preceding section.

THEOREM. *Every perfectly compact set in a topological space is compact.*

Let E be a perfectly compact set in a topological space and let A be any infinite subset of E. Assume the elements of A arranged in an irreducible ordinal series:

$$p_0, p_1, \cdots, p_\alpha, \cdots \qquad (\alpha < \Omega), \tag{1}$$

and let

$$G_\alpha = \sum_{\beta \geq \alpha} p_\beta.$$

The sequence $\mathfrak{S} = [G_\alpha | \alpha < \Omega]$ is decreasing and $\Pi G_\alpha = 0$. By hypothesis the

sequence $\mathfrak{S}$ is closed in the space (P, K), that is, there is a point p common to the sets $K(G_\alpha)$, $\alpha<\Omega$. Since $G_0=A$, p belongs to $K(A)$. Hence E is compact.

THEOREM. *If a set E in a topological space is perfectly compact every infinite subset of E of regular cardinal number has a nuclear point.*

Let H be a subset of E of regular cardinal number and let Ω be the least ordinal corresponding to the cardinal $|H|$. Let a series of the type (1) above denote a representation of H as an ordinal series. The corresponding decreasing sequence $\mathfrak{S}$ is closed in E, that is, there is a point q of E which is common to all the sets $K(G_\alpha)$.

Let V be any neighborhood of the point q. There is an ordinal α_0 such that for all $\alpha>\alpha_0$, G_α does not contain q. For each $\alpha>\alpha_0$ there is a point q_α of G_α in V. Let Q be the set of all distinct points q_α so defined. For a given element q_α of Q let β denote the index such that q_α is in G_β but not in $G_{\beta+1}$. The index β is in fact that index α which corresponds to q_α regarded as an element of H and determined by the correspondence (1). These indices β are such that for every α there is a $\beta>\alpha$. Since the cardinal $|H|$ is regular it follows that $|Q|=|H|$. Therefore the point q is a nuclear point of the set H.

Sierpinski has shown by an example for the case of spaces (P, V) that this theorem cannot be extended to sets of irregular cardinal number.*

THEOREM. *If every infinite subset H of regular cardinal number of a set E in a topological space has a nuclear point and $\mathfrak{S}=[G]$ is an open decreasing sequence of subsets of E, then there is a point common to the sets $L(G)$.*

This theorem can be deduced from the results of §16. It is sufficient to state that a point p is in the relation T to a set V if V contains p.

Since every space (P, L) is a space (P, V) it will be seen that we have obtained a generalization of a theorem which I demonstrated in an earlier article for sets of regular cardinal number.†

THEOREM. *A necessary and sufficient condition that a set E in a space (P, L) derived from a topological space be perfectly compact is that every infinite subset of E of regular cardinal number have a nuclear point.*

The following example of a non-monotonic set-function $K(E)$ shows that this theorem admits no further generalization.

22. **Example of a space (P, K) which is nuclear but not perfectly compact.** Let P be the class of all positive integers. Let

$$S_1 = p_1, p_2, p_3, \cdots \tag{1}$$

* Bulletin of the American Mathematical Society, vol. 32 (1926), pp. 649–653.

† Bulletin of the American Mathematical Society, vol. 30 (1924), p. 511.

be the sequence of all prime numbers. Let

$$S_n = p_n p_{n+1},\ p_{n+1}p_{n+2},\ p_n p_{n+2},\ \cdots . \tag{2}$$

Let

$$G_n = S_n + S_{n+1} + S_{n+2} + \cdots .$$

The function $K(E)$ is defined as follows. If E is finite, $K(E)=0$. If E is infinite and n is the least integer for which E contains all but a finite number of elements of S_n, $K(E)=p_n$. If E is infinite and does not satisfy this condition, $K(E)=1$.

The sequence $G_1, G_2, G_3, \cdots$ is decreasing and $K(G_n)=p_n$. Thus the space is not perfectly compact, although it is perfectly L-compact.

The space is nuclear. Let q be a K-point of a set E, and let Q be any set containing only a finite number of points of E. Then by definition, q is a K-point of $E-EQ$ and cannot be interior to Q. Thus Q is not a neighborhood of the element q. The function $K(E)$ is obviously not monotonic. This example may be extended to include the case of nuclearity of any order.

23. **Perfectly self-compact sets in a topological space.** The following theorem generalizes a result of Sierpinski.* The method of proof was suggested by W. L. Ayres.

THEOREM. *Every perfectly self-compact set E contains a nuclear point of itself.*

Suppose the theorem is not true. Then for every point p of E there exists a neighborhood V_p such that

$$|V_p \cdot E| < |E| .$$

The family $\mathfrak{F}$ of neighborhoods V_p so defined is a proper covering of E. Let

$$\mathfrak{F}_1 = [V_0, V_1, \cdots, V_\alpha, \cdots, |\ \alpha < \Omega_\mu],$$

where $\mu \leqq |E|$, be the corresponding normal subfamily, and let $Q=[q_\alpha]$ be the corresponding associated set.

Consider the decreasing sequence

$$G_0 = E, G_1 = E - EV_0, \cdots, G_\alpha = E - E\cdot\Big(\sum_{\beta<\alpha} V_\beta\Big).$$

Since E is perfectly self-compact, there exists an element p of E which, for every α, belongs to $K(G_\alpha)$. But p is interior to some set V_{α_0} of $\mathfrak{F}_1$. And if $\beta_0 > \alpha_0$, then $G_{\beta_0} = E - E\cdot(\sum_{\beta<\beta_0} V_\beta)$ has no point in common with V_{α_0}. Hence p does not belong to $K(G_{\beta_0})$, the desired contradiction.

* Bulletin of the American Mathematical Society, loc. cit.

III. Properties of the class of all continuous functions on a topological space

24. Continuous functions on a topological space. Let $M(f, V)$, $m(f, V)$ denote the upper and lower bounds of a real valued single valued finite function $f=f(p)$ on a set V. Then

$$w(f,V) = M(f,V)$$

is the oscillation of the function f on the set V if these bounds are finite. If one or both of the bounds are infinite the oscillation is infinite.

If the oscillation of a function f has the minimum zero on the family of all possible neighborhoods of a point p, the function is continuous at p.* A function is continuous on a set E if it is continuous at every point of E. Continuity relative to a set E is defined by means of the oscillation on the set EV.

Denote by A_f the set of all functional values of a function $f(p)$. Let M, m (finite or infinite) be the upper and lower bounds of A_f.

The notation $E(f<a)$ refers to the set of all points at which the value of the function f is less than a. The meaning of the notations $E(f\leqq a)$, etc., is evident.

For every continuous function $f(p)$ on a topological space, the sets $E(f<a)$, $E(f>a)$ are *open*, and the sets $E(f\leqq a)$, $E(f\geqq a)$ are completely closed.

25. Continuous functions on a neighborhood space. It is important to observe that the preceding definition of continuous function is a neighborhood definition. It follows readily that if (P, V) is the neighborhood space or class (V) derived from a topological space (P, K), then the class of all continuous functions is the same for both spaces. This implies that under this definition the theory of continuous functions attains its greatest generality in the classes (V).

26. Linear series of closed sets. Let $f(p)$ be any function on a class P. The set $L_x=E(f=x)$, of all points p such that $f(p)=x$, will be called the x-level of the function f. If $x\neq x'$ then $L_xL_{x'}=0$. If the function f is continuous, the set L_x is completely closed. Furthermore, if we set $m_f=a$, $M_f=b$, we obtain for each value of x in the interval $a\leqq x\leqq b$ a set L_x (which may be null) and the set

$$F_x = \sum_a^x L_x$$

is completely closed for each value of x. A series of closed sets of this type will be called a linear series.

* Fréchet II, p. 363.

A necessary and sufficient condition that a linear series $L_x(a \leqq x \leqq b)$ correspond to a continuous function $f(p)$ on a topological space (P, K) is that the set

$$\sum_{x=x'}^{x=x''} L_x$$

be completely closed for every pair of values of x', x''.

27. **Conditions for the existence of non-constant continuous functions.** It is evident that the theory of continuous functions on a space is barren in case all continuous functions are constant. This is true, for example, for all enumerable connected spaces. Fréchet has proposed, in correspondence with Urysohn and myself, the following problem: *characterize those spaces in which there exists a non-constant continuous function.* This problem has been solved for the spaces of Hausdorff by Urysohn.† His solution is easily extended to topological spaces in general.

A family of open sets G is *normal* provided

(a) there exist two sets G_0, G_1 of the family which are distinct and such that $G_0 < G_1$;

(b) if G_0, G_1 are any two sets of the family such that $G_0^* \leqq G_1$,‡ there is a set G such that

$$G_0^* \leqq G \leqq G^* \leqq G_1.$$

THEOREM (Urysohn). *A necessary and sufficient condition that a topological space admit the existence of a non-constant continuous function is that it contain a normal family of open sets.*

As the proof of this proposition given by Urysohn for Hausdorff spaces can be extended without serious changes to the general case the demonstration is omitted.

It should be noted that the requirement of normality is significant only in connected spaces. The preceding theorem has the following corollary:

THEOREM (Urysohn). *Every connected topological space which contains a normal family of open sets has the cardinal number of the continuum at least.*§

---

† *Über die Mächtigkeit der zusammenhängenden Mengen*, Mathematische Annalen, vol. 94 (1925), p. 290.

‡ The set G^* is the least completely closed set containing G. See §8 above.

§ Urysohn (loc. cit.) has given an example of an enumerable connected set on which every continuous function is constant. The following space illustrates this possibility very simply. Let P be the class of all positive integers. For each integer p the mth neighborhood $V_{p,m}$ is the set $[p; m+1, m+2, \cdots]$.

28. Another form of the problem of the existence of non-constant continuous functions. Fréchet has proposed the problem of the existence of non-constant continuous functions in the following form: *characterize the spaces such that for every subset E there is a continuous function on the space which is not constant on E.*

In this case it is both necessary and sufficient that there exist for every two distinct points p, q a function f, continuous on P, such that $f(p) \neq f(q)$. And as in the preceding article it is necessary and sufficient that there exist for each pair of points p, q a normal family of open sets G, and in this family two sets G_0, G_1, such that

$$p < G_0 < G_1$$

and q is not an element of G_1*.

29. The neighborhood space defined by the class of all continuous functions on a topological space. In terms of the class C of all continuous functions on a topological space we define a neighborhood space (P, W) in which the set-function W is defined by the family of all possible sets

$$W = E(0 \leqq g < 1/n),$$

where g denotes a non-negative continuous function.

The space (P, W) is easily seen to be monotonic, accessible and regular.† It has the properties A and B of Hausdorff. It is at once evident that the set-functions V and W derived from a postulated set-function K will not be equivalent in general. Since every neighborhood W is a neighborhood V, the condition for the equivalence of the two systems of neighborhoods reduces to the following: every neighborhood V contains a neighborhood W. Hence

A necessary and sufficient condition that the set-functions $W(E)$, $V(E)$ derived from a set-function $K(E)$ be equivalent is that there exist for every point p and neighborhood V of p a non-negative continuous function g such that for some value of n the set $W = E(g < 1/n)$ contains p and is a subset of V.

The following theorem is equally evident.

THEOREM. *The space (P, K) is equivalent to its derived space (P, W) with respect to continuity. That is, every function which is continuous on one space is continuous on the other.*

30. Analysis of the properties of the class of all continuous functions. Certain properties of the class C of all continuous functions on a topological

† A space is regular provided there exist, for every point p and completely closed set B not containing p, a pair of disjoined open sets $U > p$, $V > B$. Since p is not a W-point of B there is a function g and an integer n such that the set $U = [0 \leqq g < 1/n]$ contains p and no point of B, while $V = [g > 1/n]$ contains B. This does not imply that every pair of points is separated by open sets.

space (P, K) hold for all topological spaces. Among these properties are the following:

The class C contains all constant functions. If a function f belongs to C and g is a constant, $f+g$ belongs to C, likewise fg.

If f is a continuous function of a real variable x and g belongs to C, then $f(g)$ belongs to C.

The following simple example shows that the sum or product of two continuous functions may be discontinuous. Let $P=[a, b, c]$ be a space of three elements. The only point of accumulation is a which is a K-point of $E=[b, c]$ only. Let $f(a)=f(b)=1$, $f(c)=2$; $g(a)=g(c)=3$, $g(b)=4$. Then f and g are continuous on P while their sum and product are both discontinuous.

It is quite easy to show that if the class C is additive it is multiplicative.

THEOREM. *If C is multiplicative it is additive.*

Suppose C is not additive at a point p. Then functions f, g exist such that

$$h = f + g$$

is discontinuous at p. We may assume that $|f| \leqq 1$ and that $g(p)=0$. Since for every e there is a neighborhood U of p on which osc $g<e$ and therefore $|g|<e$ we have, at once, fg vanishes and is continuous at p. But the function

$$(1 + f)(1 + g) = 1 + f + g + fg$$

is certainly the discontinuous product of continuous functions.

Thus the algebraic closure properties of C are of two kinds; those that condition the space and those that do not. The fundamental property of the first kind is the additive property:

(1) *The sum of any two functions of the class C is a function of class C.*

It will be noted that the statement of the property is in no way related to the property of continuity assumed for the functions of the class C. In attempting to characterize a class of functions which is the class C this property would naturally be assumed in spite of the restriction on the space (P, K) which it involves. The problem of characterizing the class C varies with the space P. For example on a space of one element all functions are continuous. In the space of the previous example the class C is characterized by the fact that there is a point $p(=a)$ such that every function of f which has the value $f(p)$ at some other point of P is a function of C. It is evident that both of these characterizations involve properties not normally possessed by the class of all continuous functions.

Many of the classic properties of the class C of all continuous functions on a closed finite linear interval are in fact properties of the set A of functional

values. Thus A is itself a closed finite interval. This statement is implied by the statement that a continuous function f is bounded, attains its bounds, and every value between. It will be shown in §§32, 33 that these three properties are consequences of the following two:

(2) *For every function f of the class C the set A_f is closed.*

(3) *For every function f of the class C the set A_f is dense on an interval.*

The separation of the points of space is implied by the property

(4) *If p and q are any two distinct points there is a function f of C such that $f(p) \neq f(q)$.*

The class C should possess the fundamental property of Weierstrass:

(5) *There exists in C a finite or enumerably infinite family G of linearly independent functions*

$$g_1, g_2, \cdots, g_n, \cdots$$

such that each function f of C is the limit of a uniformly convergent sequence of functions of G.

It will be shown that the five properties so far stated are independent. The question of their complete independence has not been attacked.

The property of uniform continuity is easily defined for metric spaces in terms of the distance between two points. Since the distance $f(p) = \delta(p, q)$ between a fixed point p and a variable point q is a continuous function of p it is evident that uniform continuity is essentially a relationship between the class C and a subclass of C. The property of uniform continuity thus takes the following form:

(6) *There exists in C a family of functions $u(p)$, with the following properties:*

(a) *if p is a point there is a unique function u_p associated with p;*

(b) *if f is a function of C and e is a positive number there is a number $d(>0)$ such that for each point p and corresponding function u, $|f(p)-f(q)|<e$ whenever $|u_p(p)-u_p(q)|<d$.*

If this property is present the functions of C will be said to be uniformly continuous relative to the family $[u]$. The relation between this property and its predecessors has not been exactly determined. It is however quite easy to see that a class C with this property is also (1) additive.

In §29 we called attention to the derived space (P, W) defined by the class C and found the condition that this space be equivalent to the space (P, V) derived from a postulated space (P, K). This condition is the seventh of our list of fundamental properties. It is the first to utilize explicitly the given limiting relations.

(7) *For every point p and neighborhood V of p such that $C(V) = P - V$ is not null, there is a function g of C which is bounded from $g(p)$ on $C(V)$.*

Here the function g is dependent on the neighborhood V. In case it is uniformly independent of V we obtain a property denoted by (7′).

Since the property (7) implies the equivalence of the spaces (P, K) and (P, W), all properties of the latter space become properties of the space (P, V) in the presence of (7).

The property of points expressed in (7) provides a basis of classification. A point for which this condition fails involves an interesting singularity of the function $K(E)$.

If the class C has this property (7) the space (P, V) must be regular and accessible.

It is easily shown by extending the proof of a theorem of Urysohn* to topological spaces in general that if the space is normal† the following property is present:

(8) *If A and B are any two completely closed disjoined sets there exists a function f of C, $0 \leqq f \leqq 1$, such that $f=0$ on A, $f=1$ on B.*

The following sections are devoted to a detailed study of the relationships between these eight fundamental properties of the class C and the corresponding properties of the underlying space (P, K). The following relationships are quite easily established. The first property is implied by the sixth and seventh respectively, the seventh is a consequence of the eighth.

31. Spaces for which the class C is additive. If the class C is additive it must be additive at each point of space. If a point p is such that every continuous function is constant on some neighborhood V of p, the class C is additive at P. Points of this character will be called level points of space. A *level* L of a space P is a set such that every continuous function is constant on L. The levels of a space P are completely closed sets.‡

If a point p is not a *level* point, there is for every neighborhood V of p a continuous function f which is not constant on V. From the result of §27 we see that the function f determines a normal family of open sets V each containing p. A point which is not a level point will therefore be called a *normal* point. It is evidently a regular point.

---

* Loc. cit. See also E. W. Chittenden, Bulletin of the American Mathematical Society, vol. 33 (1927), pp. 17–18.

† A space is normal in case for every two completely closed disjoined sets A, B, there exist disjoined open sets V, U such that $A \leqq V$, $B \leqq U$.

‡ The following example shows that a level need not be connected. Let P be the set of all positive integers and let the points 1, 2 be points of accumulation of every infinite subset of P. Then the disconnected set 1, 2 is a level of P. It is, however, necessary that P be connected relative to a level L. That is, if $P=A+B$ where $H(A, B)=0$, then one of the sets A, B must contain L. For any point p the associated level L is the set common to all the sets $E(f=f(p))$ for a continuous f.

THEOREM. *The points of a topological space fall into the two classes, level points and normal points.*

A point p at which the class C is not additive is normal. There must exist continuous functions f, g, such that their sum $f+g$ is discontinuous at p. It follows from the definition of continuity that the minimum of the oscillation of $f+g$ on the neighborhoods of p is positive. Since open neighborhoods V, W of p exist on which the oscillations of f and g respectively are arbitrarily small it follows that in general the product VW will not be a neighborhood of p. The following theorem can be easily inferred.

THEOREM. *A necessary and sufficient condition that the class C be additive at a normal point p is that there exist two normal families of open sets containing p such that if U is a set of one family and W a set of the other, then the set UW is an open set.*

The following theorem is now evident.

THEOREM. *In a topological space satisfying the second axiom of Hausdorff, that is, the condition that the product of open sets is an open set, the class C is additive.*

In particular, if the neighborhoods of each point can be defined in terms of an écart $\delta(p, q)$, this condition is satisfied. From §36 we see that property (6) implies property (1).

THEOREM. *If the class C is* (1) *additive the associated derived space* (P, W) *has the second property of Hausdorff, the function* $W(E)$ *has the fundamental properties* I, II, III, *and the space* (P, W) *is both accessible and regular.*

32. Spaces for which every continuous function is bounded. It is easy to see that if every continuous function is bounded, every continuous function attains its bounds. For if a real number a is a bound of a continuous function f and is not attained it must be a limit of functional values. Hence the function $1/(f-a)$ is both continuous and unbounded. Likewise the set A_f of functional values of a continuous function f must be closed. Conversely, if the set A_f is closed for every continuous function f, every continuous function must be bounded. Otherwise for some continuous function f, the continuous function $1/(1+|f|)$ would fail to attain its lower bound zero.

THEOREM. *A necessary and sufficient condition that every continuous function on a topological space be bounded is* (2): *the set of functional values of every continuous function is a closed set.*

If a topological space admits an unbounded continuous function f there exists a corresponding normal series of open sets (coterminal with a series of

type ω) which have no common element. Conversely the existence of a normal series of this type permits the definition, by an obvious modification of the method of Urysohn referred to in §27, of an unbounded continuous function.

THEOREM. *A necessary and sufficient condition that for every function f of the class C of all continuous functions the set A_f be limited or closed is that every normal series of open sets which is coterminal with a series of type ω determine at least one point common to the sets of the series.*

If the space (P, V) derived from a topological space is compact every continuous function is bounded. But there exist non-compact spaces on which every continuous function is bounded, for example, in case a fixed point p_0 is an element of every neighborhood of every point.

33. Spaces for which the set of functional values of every continuous function is dense on an interval. Suppose that the set A_f is dense on an interval I for some continuous function f, but that A_f is not a continuum. Then $A_f = A_1 + A_2$, where neither of the sets A_1 or A_2 contains a point or limit point of the other. Let P_1, P_2 be the subsets of P on which the functional values of f lie in A_1, A_2 respectively. The function equal to 0 on p_1, 1 on P_2 is continuous, and its set of functional values is not dense on an interval.

THEOREM. *A necessary and sufficient condition that the set A_f be dense on an interval for every continuous function (f) is that the space be connected. Under either hypothesis every continuous function attains all values between its bounds.*

34. The property (4). In this section we consider spaces for which the class C has the property

(4) *If p and q are distinct points there is a function f of C such that $f(p) \neq f(q)$.**

THEOREM. *A necessary and sufficient condition that the class C on a topological space have the property (4) is that for every pair of points p, q there is a normal series of open sets containing p of which q is not a common element.*

If a single function is effective for all pairs of points the space is homeomorphic with a subset of the linear continuum. The property (4) is satisfied vacuously if there is but one point in space. It is present in any space which is disconnected between every pair of points.

* Because of the close relationship between properties (4) and (7) the following example is of interest. Let $P = [p + \{p_n\} + \{\{p_{mn}\}\}]$. Let p be the sequential limit of p_n, and p_n the sequential limit of p_{mn}. There are no other limit relations in P. No subset of the set $E = \{\{p_{nm}\}\}$ has p for a point of accumulation. Hence the set $V = [p + \{p_n\}]$ is a neighborhood of p. But the values assumed on E by any continuous function on P must have $f(p)$ for a limiting value. In this case the class C has property (4) but not property (7).

The Hausdorff property (D) (for every two points p, q there exist disjoined neighborhoods V_p, V_q) is implied by property (4). The third and fourth conditions of F. Riesz, a finite set has no point of accumulation, and a point is uniquely determined by the sets of which it is a point of accumulation, are satisfied in the derived space (P, V) when property (4) is present.

If the function f of property (4) is independent of the point q we obtain a property (4′) which appears to be topologically independent of (4), although the exact nature of the difference is not evident. If the space is compact and has property (4′), it is easy to show that it has the properties (A), (B), (D) of Hausdorff and also satisfies the first axiom of enumerability.

35. **The fifth property.** This property is suggested by the Weierstrassian theorem that a continuous function of a real variable is the limit of a uniformly convergent sequence of polynomials. This property is present in every space which can be represented as the sum of a finite number of distinct levels.

Let C_0 be the family of functions

$$g_1, g_2, g_3, \cdots.$$

Let b_n be the lower bound of g_n and set

$$Q_{nk} = E[b_n \leqq |g_n| < b_n + 1/k].$$

The family of all sets Q_{nk} obtained by varying the integers n and k is enumerable and equivalent to the family of neighborhoods W determined by the class C. In fact, if g is any continuous function, and $W = E[|g - g(p)| < a]$, we may, because of the uniform convergence of a sequence of functions g_{n_m} of C_0 to the continuous function $f = g - g(p)$, find an integer m for which $|g_{n_m} - f| < a/2$. Consequently the points of the set $Q_{n_m k}$ on which $|g_{n_m}| < a/2$ belong to the set E. Thus every neighborhood W contains a set Q_{n_m}. The sets Q_{n_m} are by definition open sets. It follows that the space (P, W) is perfectly separable.

THEOREM. *If the class C has the property* (5) *the corresponding space* (P, W) *is perfectly separable.*

That the Weierstrassian property (5) is present in any separable metric space follows readily from the theorem of Urysohn that every such space is homeomorphic to a subset of the Hilbert space.*

36. **On uniform continuity relative to a family of continuous functions. The sixth property.** In the sixth fundamental property of the class C we assume the existence of a family of continuous functions u_p, one for each point

* Mathematische Annalen, vol. 92 (1924), p. 302.

of space, such that for every continuous function f and positive number e there is a number d such that if $|u_p(p)-u_p(q)|<d$ then $|f(p)-f(q)|<e$. It will be convenient to introduce the distance function

$$\delta(p,q) = |u_p(p) - u_p(q)|$$

which is continuous in q for fixed values of p.* The sixth property is possessed by all spaces which are composed entirely of isolated elements, since in this case all functions are continuous, and also uniformly continuous relative to an arbitrary class of functions u_p. If a space of this type has infinitely many points, we obtain unbounded uniformly continuous functions.

The space (P, δ) defined by this distance function $\delta(p, q)$ is equivalent to the space (P, W) determined by the entire class C. For since u_p is a continuous function its oscillations on the neighborhoods W of p have zero for their minimum. Likewise the oscillation of a continuous function f on the neighborhood $V_n=[\delta(p, q)<1/n]$ has the minimum zero because of uniform continuity.

THEOREM. *If the class C has the property* (6) *it is* (1) *additive and there is an écart* $\delta(p, q)$, *continuous in q, such that the spaces* (P, δ) *and* (P, W) *are equivalent.*

37. **The seventh property.** Consider the property (7): for every point p and neighborhood V of p such that the complementary set $C(V)=P-V$ is not null, there is a continuous function f which vanishes at p and is bounded from zero on $C(V)$.

It is shown in §29 that this property is the condition that the spaces (P, V) and (P, W), where the set function W is defined by the class C, be equivalent. Thus a space (P, V) in which (7) is satisfied is required to be both accessible and regular.

A necessary and sufficient condition that a space with the property (7) be a regular Hausdorff space is that it satisfy the first axiom of separation, that is, that every pair of distinct points lie in disjoined open sets, and that the class C be additive.

In particular if the class C has the properties (1), (4), (7), the space (P, V) is a regular Hausdorff space.

From properties (6), (7), we may conclude that the space (P, δ) of §36

* In general the function $\delta(p, q)$ is unsymmetric. It does not appear to have been previously observed that the assumption of a *uniformly continuous* symmetric écart $\delta(p, q)$ is essentially equivalent to the assumption of the axiom of the triangle, $\delta(p, q)\leqq\delta(p, r)+\delta(r, q)$. Note added during correction of proof: This proposition was established by Menger, Mathematische Annalen, vol. 100, Nos. 1-2, August 28, 1928, p. 145.

is equivalent to the space (P, W) of §29. The results of this section are summarized in the following theorem.

THEOREM. *If in a space (P, V) the class C of all continuous functions has the property (7), the space is accessible and regular. If properties (1), (4) are present the space is regular Hausdorff space. Properties (5), (7) together ensure that the space be accessible, regular and perfectly separable (therefore normal). Properties (6) and (7) together imply that the space (P, W) is equivalent to a space (P, δ) where $\delta(p, q)$ is continuous in q.*

38. **A stronger property.** We consider some consequences of the presence of the property (7′) obtained by making the function g of (7) independent of the neighborhoods of p.

(7′) *For any point p there is a function g which is bounded from $g(p)$ on $C(V)$ for every neighborhood V of p whose complement is non-null.*

Since (7′) implies (7) it follows at once that the space is regular and accessible. But it is quite easy to see that the product of any two open sets is an open set.

This property permits a statement of the converse of the proposition, every continuous function on a compact space is bounded. It is possible to show that if a space (P, V) with the property (7′) is not compact there exists an unbounded continuous function.

Since the space is not compact there exists an infinite set A of points with no point of accumulation. For any point p the set $V = P - A - p$ is a neighborhood, and $C(V) = A - p$ is not null. Then by hypothesis there is a function which vanishes at p and at no other point. Let $Q = [q_n]$ be an enumerable set which has no V-point. Then for each point q_n we have a continuous function g_n, $0 \leqq g_n(q_n) \leqq 1$, which vanishes at q_n and at no other point. Let us choose a closed neighborhood $\overline{V}_m$ which contains no point of $Q - q_m$. On $C(\overline{V}_m)$, $g_m \geqq d_m > 0$. It is possible by a well known process to define integers k_m so that the infinite product

$$f_n = \prod_{m=n} g_m e^{\gamma_m},$$

where

$$\gamma_m = 1 - g_m + (1 - g_m)^2/2 + \cdots + (1 - g_m)^{k_m}/k_m,$$

converges uniformly in some neighborhood of every point p. It is sufficient to choose k_m equal to the larger of $1/d_m$ and $m\,|(\log 2)/\log(1-d_m)|$. The function f_n so defined is continuous, vanishes at q_n, $q_{n+1}, \cdots$, and at no other points. The function $h_m = f_n$, if $f_n \leqq 1$, and $= 1$, if $f_n = 1$, is continuous. The function

$$h = \sum_1^\infty \frac{1}{2^n} h_n$$

is continuous, does not vanish, and satisfies the inequality

$$h(q_n) = \frac{1}{2^{n+1}}h_{n+1}(q_n) + \frac{1}{2^{n+2}}h_{n+2}(q_n) + \cdots \leqq \frac{1}{2^n}.$$

Therefore the continuous function $1/h$ is unbounded.

THEOREM. *In a space with property* (7′) *a necessary and sufficient condition that* (2) *every continuous function be bounded, is that the space be compact.*

It would be of interest in this connection to determine the most general topological space in which compactness is equivalent to property (2).

THEOREM. *If the class C of all continuous functions on a neighborhood space* (P, V) *has the property* (7′), *then the space is regular and satisfies the Hausdorff axioms* (A), (B), (C). *The class C is additive. If every continuous function is bounded the space is compact.*

39. **The property of Hahn.** The eighth of the series of properties discussed may be called the property of Hahn because it appears as part of Hahn's proof of the existence of non-constant continuous functions in a space admitting a uniformly regular écart (originally called "voisinage" by Fréchet). The class C on a space (P, V) has the property (8) *if for every pair of disjoined closed sets, A, B, there exists a continuous function* f, $0 \leqq f \leqq 1$, *which vanishes on A and is equal to* 1 *on B.*

This property stated for (P, V) is equivalent to the property defined for spaces (P, K) by replacing the word closed by "completely closed." Properties (7) and (7′) are immediate consequences of (8). The importance of this property can be inferred from the following theorem which is essentially due to A. D. Pitcher:

THEOREM (Pitcher). *If the class C has the properties* (2), (4), (6), (8), *the space* (P, K) *is a compact (separable) metric space. The class C has the additional properties* (1), (2), (5), (7), (7′).

This theorem follows from the fact that (8) implies (7), the preceding theorem, a theorem of A. D. Pitcher,* the equivalence of "distance" and "regular écart," the well known fact that every compact metric space is separable,† and the fact that all the properties mentioned are present if P is a metric space.

40. **Independence.** The following list of examples shows that the properties (1)-(5) of the class C are independent. That property (7) is indepen-

* A. D. Pitcher, loc. cit., p. 76, Theorem 19.

† Fréchet, Palermo Rendiconti, vol. 30 (1910), p. 1–26.

dent of (1)-(6) is also established. Similarly (7′) is independent of (1)-(7). We have no example showing that (8) is independent of the remaining eight properties, and there are reasons for thinking it is not. While properties (2) and (6) are independent, as is shown by examples (5) and (8), we have no example showing that (6) is independent of (1)-(5). It would be interesting to determine the exact relationship between boundedness and uniform continuity.

In the following examples the properties not present are indicated. The proofs will be left to the reader.

1. (1), (6), (7), (8). Let $P=[0\leqq p\leqq 1; 2]$. Limit is defined as usual for the points of the interval, $0\leqq p\leqq 1$. Let $p_{2n}=1/n$, $p_{2n+1}=1-1/n (n=1, 2, 3, \cdots)$. The neighborhoods of the point 2 are $V_{2n}=[2; p_{2n}, p_{2n+1}, \cdots]$, $V_{2n+1}=[2; p_{2n+1}, p_{2n+1}, \cdots]$.

2. (2), (6). The class P is the segment $0<x<1$, with limit defined as usual.

We have no example of a space with all of the properties except (2).

3. (3). P is composed of two isolated points.

4. (4). P is the interval $0\leqq p\leqq 1$, together with the point $p=2$. Limit is as usual, except that the point $p=2$ is in every neighborhood of $p=0$.

5. (5), (6). The class P is the space defined by Hausdorff* composed of a non-enumerable well ordered set with the interstices filled by linear continua.

We have no examples of the class (5) or (6).†

6. (7), (7′), (8). The class P is the interval $0\leqq p\leqq 1$, limit is defined as usual except that the point $p=0$ has its neighborhoods defined by the formulas [all irrational numbers $p<1/n$], [all rational numbers $p<1/n$] $(n=1, 2, 3, \cdots)$.

7. (7′), (8). Let P be the interval $0\leqq p\leqq 1$. The neighborhoods of all points $p>0$ are as usual. From each neighborhood of zero remove its middle point. Then the space P is not compact, for the sequence of points $1, 1/2, 1/3, \cdots$ has no point of accumulation. The class C of all continuous functions is unaffected by this modification of the neighborhoods of zero.

8. (2), (3). The class P is the class of positive integers, with each point isolated.

---

* Tietze, Mathematische Annalen, vol. 91 (1924), p. 218, Example B_8.

† Examples 2 and 5 show the independence of properties (2) and (6).

UNIVERSITY OF IOWA,
IOWA CITY, IOWA

PROPERTIES OF FUNCTIONS REPRESENTED BY THE DIRICHLET SERIES $\sum(a\nu+b)^{-s}$, OR BY LINEAR COMBINATIONS OF SUCH SERIES*

BY

J. I. HUTCHINSON

Introduction. The following paper deals with the functions

$$Z(a, b, s) = Z(s) = \sum_{\nu=0}^{\infty}(a\nu + b)^{-s}, \tag{1}$$

in which a and $b(\leqq a)$ represent real, positive numbers, and $s=\sigma+it$. They include as a special case $(a=b=1)$ the Riemann function $\zeta(s)$, characterized by Gram as "une des plus remarquables acquisitions de l'analyse moderne"; while simple linear combinations of $Z(a, b, s)$ for $b=1, 2, \cdots, a-1$, when a is a fixed integer, give the Dirichlet L-functions. All of these special functions, particularly $\zeta(s)$, are of the highest importance in the modern analytic theory of numbers and an immense amount of investigation has been lavished upon them.

Moreover, since $Z(s)=a^{-s}\zeta(s, b/a)$, the close relationship of $Z(s)$ to the generalized ζ-function $\zeta(s, w)$† lends an additional importance to these functions. Accordingly, §§1–8 may be regarded as a contribution to the theory of $\zeta(s, w)$. The function $Z(s)$, however, is more convenient for the purposes of the present paper.

The great value of the above mentioned and other functions for the theory of numbers has stimulated much activity during the past twenty years in the general theory of functions represented by Dirichlet series, quite apart from any question of applications, although, no doubt, with the idea of throwing more light on the properties of functions having such use. This is the point of view in what follows.

The main object of §§1–8 is to determine the real zeros of $Z(s)$ (formula (7))‡ and, approximately, the region of the plane in which the complex zeros are located (§4).

* Presented to the Society, September 7, 1928; received by the editors in May, 1928.

† For $\zeta(s, w)$ see Bohr and Cramér, *Die neuere Entwicklung der analytischen Zahlentheorie*, Encyclopädie der Mathematischen Wissenschaften, II C 8, 1922, pp. 777–779. Recently $\zeta(s, w)$ has been adopted into the analytic theory of numbers as a basis of comparison. See E. Landau, *Vorlesungen über Zahlentheorie*, vol. 2, 1927, p. 9 ff.

‡ These are in striking contrast to the real zeros of $\zeta(s)$ which occur at $s=-2, -4, \cdots, -2n, \cdots$, so that $\zeta(s)$ contains the factor $1/\Gamma(1+s/2)$. For this reason these are called the "trivial zeros" of $\zeta(s)$. The L-functions, likewise, have trivial zeros on the negative real axis.

That complex zeros do occur is established (§5) by the actual calculation of these zeros for $Z(3, 1, s)$ and $Z(3, 2, s)$ between the limits $0<t<50$. The interesting fact is revealed in these two cases that the complex zeros (as far as determined) are scattered about irregularly in the narrow strip $0<\sigma<1$ and do not lie on a straight line as is presumably the case with the ζ- and L-functions.*

In §6 the formulas for numerical computation are developed and discussed, and in §7 the curves $C=0$, $S=0$, along which the real and imaginary parts of $Z(s)=C(\sigma, t)+iS(\sigma, t)$ vanish, are investigated. It is found that all these curves, excepting the curves C for which $b=1$, have a series of right-hand ($\sigma>0$) asymptotes occurring at equal distances, a result which is very useful in numerical calculation and in the determination of the number of imaginary zeros in the region $0<t<T$(§16). The behavior of these curves for $\sigma<0$ is much more complicated, the approximate variations of C and S being given by (18), and of $Z(s)$ by (19).

The remainder of the paper is devoted to the problem of determining linear combinations of $Z(a, b, s)$ which satisfy functional equations of a type similar to that of which $\zeta(s)$ is a solution. This problem, which was solved by Cahen† for a a prime, is completed for all integer values of a (Theorems 3 and 4). The functions $f(s)$ thus determined include the L-functions as very particular cases and form a large class of new functions which seem to be worthy of further study. The problem of functional equations is somewhat generalized (§18) so as to include the L-functions with imaginary characters.

In §15–17 we deal very briefly with the complex roots of the functions $f(s)$, the number of such roots being infinite (§17). It can be proved that

---

* The famous and as yet unproved "Riemann Hypothesis" (1859) supposes that all the complex zeros of $\zeta(s)$ lie on the line $\sigma=\frac{1}{2}$. Confidence in this hypothesis was first put on a solid basis by the aid of numerical computation. In particular, J. P. Gram, *Note sur les zéros de la fonction $\zeta(s)$ de Riemann*, Acta Mathematica, vol. 27 (1903), pp. 289–304, calculated the roots on $\sigma=\frac{1}{2}$ as far as $t=65$; and R. J. Backlund, *Ueber die Nullstellen der Riemannschen Zetafunktion*, Dissertation, Helsingfors, 1916, located within narrow limits all the remaining zeros up to $t=200$ and proved that all the roots within the limits $0<t<200$ lie on $\sigma=\frac{1}{2}$. Bohr and Cramér (loc. cit., p. 773, foot-note 133) remark that "this result is one of the most powerful arguments for the belief in the Riemann hypothesis." More recently the proof of this hypothesis has been extended as far as $t=300$ by J. I. Hutchinson, *On the roots of the Riemann zeta function*, these Transactions, vol. 27 (1925), pp. 49–60. Similar results have been obtained for some of the L-functions by Grossmann. The importance of this question is testified to by the increasing number of theorems in the theory of numbers whose proofs depend on an assumption of the Riemann hypothesis. See, for example, the long paper by Hardy and Littlewood quoted in the first foot-note to p. 343.

† See footnote, p. 329.

some of these functions have an infinity of roots on the line $\sigma=\frac{1}{2}$.*

1. **Fundamental formulas.** The series (1) is absolutely convergent for all values of s satisfying the inequality

$$\sigma > 1 . \tag{2}$$

To convert (1) into a series valid over a larger part of the s-plane, we start with the formula

$$Z(s) = \sum_{\nu=0}^{n-1}(a\nu + b)^{-s} + \frac{1}{2(an+b)^s} + \frac{(an+b)^{1-s}}{a(s-1)} - as\int_n^\infty \frac{\bar{\phi}_1(x)dx}{(ax+b)^{s+1}} \tag{3}$$

in which $\bar{\phi}_1(x)=x-[x]-\frac{1}{2}$; that is, $\bar{\phi}_1(x)$ is a periodic function, with period 1, which coincides with the values of $x-\frac{1}{2}$ in the interval $0\leqq x<1$.

To prove (3), denote the right member by $F(s, n)$. Separate the integral into $\int_n^{n+1}+\int_{n+1}^\infty$ and substitute $x=\xi+n$ in the first of these, giving

$$\int_0^1 \frac{(\xi-\frac{1}{2})d\xi}{(a\xi+an+b)^{s+1}},$$

which may be integrated by parts. After substituting the result in the right member of (3) we obtain

$$F(s, n) = F(s, n+1) = F(s, n+2) = \cdots = \lim_{n\to\infty} F(s, n) = Z(s).$$

In taking the limit we must restrict s by (2); but having established the relation for $\sigma>1$ we extend it to other values of σ by the principle of analytic continuation.

Formula (3) shows that $Z(s)$ has a pole of the first order at $s=1$ with residue $1/a$. By successive integration of the last term by parts we obtain the more general formula†

$$Z(s) = \sum_{\nu=0}^{n-1}\frac{1}{(a\nu+b)^s} + \frac{1}{2(an+b)^s} + \frac{(an+b)^{1-s}}{a(s-1)} + \sum_{\nu=1}^{k} T_\nu + R_k, \tag{4}$$

---

* This was first proved for $\zeta(s)$ by G. H. Hardy, *Sur les zéros de la fonction $\zeta(s)$ de Riemann*, Comptes Rendus, vol. 158 (1914), pp. 1012. Hardy's theorem was extended to the L-functions by E. Landau, *Ueber die Hardysche Entdeckung unendlich vieler Nullstellen der ζ-Funktion mit reellem Teil $\frac{1}{2}$*, Mathematische Annalen, vol. 76 (1915), pp. 212–243.

† This formula (excepting the inequality for R_k) is given by J. Grossmann, *Ueber die Nullstellen der Riemannschen ζ-Funktion und der Dirichletschen L-Funktionen*, Dissertation, Göttingen, 1913. The inequality is derived by a method similar to that employed for the ζ-function by E. Lindelöf, *Quelques applications d'une formule sommatoire générale*, Acta Societatis Scientiarum Fennicae, vol. 31 (1903), and extended by R. J. Backlund, loc. cit., p. 18. The formula for $|R_k|$ is given here for the sake of completeness but is not used in the present paper.

$$(5)\qquad T_\nu = \frac{(-1)^{\nu-1}B_\nu a^{2\nu-1}}{(2\nu)!}\cdot\frac{s(s+1)\cdots(s+2\nu-2)}{(an+b)^{s+2\nu-1}},$$

$$R_k = -\frac{a^{2k+2}s(s+1)\cdots(s+2k+1)}{(2k+2)!}\int_n^\infty \frac{\overline{P}_{2k+2}(x)dx}{(ax+b)^{s+2k+2}},$$

$\overline{P}_{2n}(x)$ being a periodic function with period 1,

$$|R_k| < \frac{|s+2k+1|}{\sigma+2k+1}|T_{k+1}|.$$

Formula (4) is valid for the half-plane $\sigma > 1-2k$.

2. **Functional equation for $Z(s)$.** By starting with the expression for $Z(s)$ as a definite integral

$$Z(s) = \frac{1}{\Gamma(s)}\int_0^\infty \frac{e^{(a-b)x}x^{s-1}dx}{e^{ax}-1},$$

and proceeding exactly as Riemann does with the ζ-function, we obtain the functional equation*

$$(6)\qquad Z(s) = \left(\frac{2\pi}{a}\right)^s \frac{\Gamma(1-s)}{\pi}\sum_{n=1}^\infty \frac{\sin\left(\frac{s}{2}+\frac{2bn}{a}\right)\pi}{n^{1-s}}.$$

3. **The real zeros of $Z(s)$.** Suppose that $s=-\rho$ (ρ a real positive number) is a zero of $Z(s)$. If we substitute $s=-\rho$ in (6), the factor $\Gamma(1-s)$ does not vanish. When ρ is sufficiently large the sign of the series is determined by its first term and this term must be numerically small, if the series is to vanish. Hence, in order that $Z(-\rho)$ may be zero it is necessary for the angle $(-\rho/2+2b/a)\pi$ to take the form $(-\frac{1}{2}\epsilon-\lambda)\pi$, ϵ being a small positive or negative number and λ an integer. From this follows

THEOREM 1. *The real roots of $Z(s)$ are given by the formula*

$$(7)\qquad -s = \rho = 2\lambda + \frac{4b}{a} + \epsilon(\lambda), \quad \epsilon(\lambda)\to 0 \quad as \quad \lambda\to\infty,$$

in which λ takes all positive integral values, including 0, and such negative integral values as will make the right member of (7) positive.

It is obvious that $Z(-\rho)$ will change sign when ρ passes through one of the values given by (7).

---

* This formula was given, for integral values of a and b, by Ad. Hurwitz, *Einige Eigenschaften der Dirichletschen Funktionen* etc., Zeitschrift für Mathematik und Physik, vol. 27 (1882), p. 86.

For sufficiently large values of λ (say $\lambda>5$), a very good approximation for $\epsilon(\lambda)$ is obtained by using the first two terms of the series in the right member of (6). Substituting $s=-2\lambda-4b/a-\epsilon$ and replacing $\sin(\epsilon\pi/2)$ in the first term by $\epsilon\pi/2$, dropping ϵ in the second term, equating the result to 0, and solving for ϵ, we obtain

$$\epsilon = \frac{\sin(2b\pi/a)}{\pi\cdot 2^{2\lambda+4b/a}}.$$

This shows that when λ increases by 1, $\epsilon(\lambda)$ is multiplied by $\frac{1}{4}$ (approximately). Hence, *as λ increases by unit increments, the function $\epsilon(\lambda)$ decreases in a geometric series (approximately and asymptotically) with $\frac{1}{4}$ as the common ratio.*

To illustrate the preceding and to give some clue to the rate of change of $\epsilon(\lambda)$ for small values of λ, the following real roots have been calculated.

Roots of $Z(3,1,s)$: $-1.401, -3.357, -5.340, -7.335, -9.334, -11.3334, -13.33336, \cdots$.

Roots of $Z(3,2,s)$: $-.535, -2.630, -4.657, -6.664, -8.666, -10.6665, -12.666625, \cdots$.

4. **The imaginary zeros of $Z(s)$.** When a and b are real, the imaginary roots of $Z(s)$ occur in conjugate pairs. In order to determine the region of the s-plane in which such zeros are situated, we note that $\Gamma(1-s)$ cannot vanish or become infinite for complex values of s. Hence the imaginary zeros of $Z(s)$ are also zeros of the function

$$f(s) = \sum_{n=1}^{\infty} \frac{\sin(s/2 + 2bn/a)\pi}{n^{1-s}}.$$

In this substitute $s=\sigma+it=1-p+it$, $p>1$, $t>0$. Then we have

$$\sin\left(\frac{s}{2}+\frac{2bn}{a}\right)\pi = \cos\pi\left(\frac{p}{2}-\frac{2bn}{a}-i\frac{t}{2}\right) = \frac{r}{2}\left(\lambda_n+\frac{1}{\lambda_n r^2}\right),$$

$$r = e^{\pi t/2}, \quad \lambda_n = \cos\theta_n + i\sin\theta_n, \quad \theta_n = \left(\frac{p}{2}-\frac{2bn}{a}\right)\pi.$$

Hence

$$f(s) = \frac{r}{2}\left(\lambda_1+\frac{1}{\lambda_1 r^2}\right)[1+\phi(s)], \quad \phi(s) = \sum_{n=2}^{\infty} \frac{\lambda_n + 1/(\lambda_n r^2)}{\lambda_1 + 1/(\lambda_1 r^2)}\cdot\frac{1}{n^{1-s}}.$$

From the relation

$$\left|\lambda_n + \frac{1}{\lambda_n r^2}\right| = (1 + 2r^{-2}\cos 2\theta_n + r^{-4})^{1/2}$$

we deduce

$$1 - r^{-2} \leqq \left| \lambda_n + \frac{1}{\lambda_n r^2} \right| \leqq 1 + r^{-2}. \tag{8}$$

From (8) follows the inequality

$$|\phi(s)| < \frac{r^2+1}{r^2-1} \sum_{n=2}^{\infty} \frac{1}{n^p} = \frac{r^2+1}{r^2-1}[\zeta(p) - 1].$$

This gives the following theorem:

THEOREM 2. *A necessary condition that $Z(s)$ may have imaginary zeros for $t>0$ is*

$$\frac{r^2+1}{r^2-1}[\zeta(p) - 1] \geqq 1, \quad p > 1, \quad r = e^{\pi t/2},$$

that is, the zero points s must be to the right of the curve

$$\zeta(1-\sigma) = \frac{2}{1+e^{-\pi t}}, \quad \sigma < 0. \tag{9}$$

In graphing (9) we assume the real axis $t=0$ to be horizontal and directed to the right. The graph starts at $t=0$, $\sigma=-\infty$ and rises slowly to $t=1$, $\sigma=-.722$. From there on it rises very rapidly and practically coincides with the line

$$\sigma = -.7143 \tag{10}$$

for t varying from 5 to $+\infty$.

In the same way we find that the critical strip for $t<0$ (the strip in the s-plane in which imaginary zeros may occur) is bounded on the left by the image of (9) below the real axis.

The right hand limit for the critical strip cannot be determined so precisely as the left hand boundary (9) and (10).

It is obvious that the right hand limit for all roots is $\sigma=\sigma_0$, in which σ_0 is not greater than the value σ_1 of σ determined by the equation

$$\sum_{\nu=1}^{\infty} (a\nu + b)^{-\sigma} = b^{-\sigma}. \tag{11}$$

Substituting $a=bp$, (11) becomes $\sum_{\nu=1}^{\infty} (p\nu+1)^{-\sigma}=1$. It is easy to see from this equation that σ_1 varies from 1 to ∞ as $p(=a/b)$ varies from ∞ to 0, or as b/a varies from 0 to ∞. When $b=a$, (11) reduces to $\zeta(\sigma)=2$, from which $\sigma_1=1.645$. Hence for $b<a$ we have $\sigma_0 \leqq \sigma_1 < 1.645$.

5. **The imaginary zeros of Z_1 and Z_2.** In order to obtain some clue to the distribution of imaginary roots, I have calculated several of them for each

of the functions $Z_1=Z(3, 1, s)$, $Z_2=Z(3, 2, s)$ whose real roots have already been determined. Formula (11) shows that all of these zeros must be to the left of the line $\sigma=1.26$, for Z_1, and to the left of the line $\sigma=1.5$ for Z_2. Denoting the coördinates of the zero points by (σ, t) the results obtained are as follows.

Roots of Z_1: (.3, 11.45), (.25, 15.2), (.6, 20.7), (.25, 23.9), (.3, 28.55), (.4, 30.6), (.5, 33.65), (.5, 37.6), (.5, 39.7), (.15, 42.25), (.35, 43.65), (.6, 47.75).

Roots of Z_2: (.15, 10.8), (.6, 16.6), (.8, 24.3), (.7, 30.8), (.37, 34.13), (.46, 37.6), (.7, 42.9), (.35, 45.34), (.1, 47.67).

6. **Numerical computation.** The imaginary zeros of functions we are dealing with are comparatively easy to calculate for values of t not too great. For this purpose we separate (4) into real and imaginary parts and denote them respectively by $C=C(\sigma, t)$, $S=S(\sigma, t)$. Introducing (for §6 only) the abbreviations $\mu=an+b$, $s=\sin(t\log\mu)$, $c=\cos(t\log\mu)$, we have

$$C=\sum_{\nu=0}^{n-1}\frac{\cos[t\log(a\nu+b)]}{(a\nu+b)^{\sigma}}+\frac{\cos(t\log\mu)}{2\mu^{\sigma}}+C_0+C_1+\cdots, \tag{12}$$

$$C_0=-\frac{\mu^{1-\sigma}[st+(1-\sigma)c]}{a[t^2+(1-\sigma)^2]}, \qquad C_1=\frac{a(st+\sigma c)}{12\mu^{1+\sigma}};$$

$$S=-\sum_{\nu=0}^{n-1}\frac{\sin[t\log(a\nu+b)]}{(a\nu+b)^{\sigma}}+\frac{\sin(t\log\mu)}{2\mu^{\sigma}}+S_0+S_1+\cdots, \tag{13}$$

$$S_0=\frac{\mu^{1-\sigma}[(1-\sigma)s-ct]}{a[t^2+(1-\sigma)^2]}, \qquad S_1=\frac{a(ct-\sigma s)}{12\mu^{1+\sigma}}.$$

If t is fairly large in comparison with σ, the terms C_2 and C_3 in (12) may be computed by the simpler approximate formulas

$$C_2=\frac{a^2}{60}\left(\frac{t}{\mu}\right)^2C_1, \qquad C_3=\frac{a^2}{42}\left(\frac{t}{\mu}\right)^2C_2, \tag{14}$$

which are equally good for S_2 and S_3 in (13).

From (5) we obtain

$$\frac{T_{k+1}}{T_k}=-\frac{a^2B_{k+1}(s+2k-1)(s+2k)}{B_k(2k+1)(2k+2)\mu^2}.$$

The factor $B_{k+1}/(B_k(2k+1)(2k+2))$ rapidly approaches the limit $1/(4\pi^2)$ as k increases. If t is not too small, a good approximation for the ratio T_{k+1}/T_k is $(at/(2\pi\mu))^2$. Consequently the terms in $\sum T_\nu$ are decreasing numerically, if $|at|<2\pi\mu$. In the computations made in connection with the

present paper and carried out with three decimals, the value of n has been chosen in (12) and (13) so that $at \leqq \mu\pi$. The terms C_ν (or S_ν), $\nu = 1, 2, 3$, then diminish very rapidly and it is unnecessary to use any other of these terms to secure a sufficient degree of approximation for the values of C and S. It is also unnecessary to calculate the remainder (a laborious process), which will be in general too small to affect materially the few decimals used. In case of doubt, the value of C (or S) has been recalculated by using an increased value of n. This can be quickly done and furnishes a good idea of the error introduced by the approximate formulas and has justified their use.

7. **The C and S curves and their right-hand asymptotes.** In calculating the roots given in §5, certain properties common to all the curves

$$C(\sigma,t) = 0, \quad S(\sigma,t) = 0 \tag{15}$$

were found very helpful. Consider first the case $b=1$. The values of C and S are readily computed by (12) and (13), the curves (15) roughly sketched, and their points of intersection located.

In order to reduce the work to a minimum, we consider what means there are of estimating where a horizontal line l, $t=k$, will intersect either curve. From (12) we observe that all the terms in C, excluding the first which is 1, diminish as σ increases and their sum becomes and remains less than 1. Accordingly, there is a point $\sigma=\sigma_0$ on the line l such that for all $\sigma>\sigma_0$, $t=k$, the function C is positive. We will speak of points on this part of l as being to the right of the curve $C=0$. I have found it most convenient (unless there are indications to the contrary) to take $\sigma=.5$ and integral values of t. Having calculated C for these values, the sign of C determines on which side of the vertical line $\sigma=.5$ the graph is situated and the numerical magnitude of C gives some rough indication (as a rule) as to how far away the graph is.

While the portion of the s-plane to the right of the curve $C=0$ is a region of positive values for the function C, it is quite otherwise with the function S whose first term is $-\sin\,[t\log(a+1)]/(a+1)^{-\sigma}$. As in the case of C, S has an invariable sign on the line l to the right of a certain point $\sigma=\sigma_0$, but when t passes through a value determined by the equation

$$t\log(a+1) = n\pi, \tag{16}$$

n an integer, the sign of S is reversed on that part of the line l to the right of the graph. In other words, *the lines* (16) *are right-hand asymptotes for infinite branches of the curve* $S=0$ *extending to* $\sigma=+\infty$.

If $b\neq 1$, *then both the curves* (15) *have an infinity of branches extending to*

$\sigma = +\infty$ in the direction of the asymptotes $t = (2n+1)\pi/(2 \log b)$ for $C=0$ and $t = n\pi/\log b$ for $S=0$.

8. **The behavior of C and S on the line $t = k, \sigma < 0$.** In equation (6) put $\sigma = -p$, $p>0$. Let $t(>0)$ have an arbitrary, fixed value and let p increase without limit. The sum in the right member of (6) may be written

$$\sin\left(\frac{\pi s}{2} + \frac{2b}{a}\right) + \epsilon(p), \quad \epsilon(p) \to 0 \quad \text{as} \quad p \to \infty .$$

Discard $\epsilon(p)$, replace the sine function by its exponential form, and drop the term containing $e^{-\pi t/2}$ which becomes very small for moderately large values of t. Writing $\Gamma(1+p-it) = \gamma_1 + i\gamma_2$, we finally obtain from (6) the approximate formula

$$Z(-p+it) = \left(\frac{2\pi}{a}\right)^{-p+it}\left(\frac{\gamma_1 + i\gamma_2}{2\pi}\right) e^{\pi t/2 + i\pi(1/2 + p/2 - 2b/a)} = C + iS.$$

From the theory of the Gamma function we have*

$$\gamma_1 + i\gamma_2 = \Gamma(1+p)e^{-P(1+p,-t)+i\omega},$$

$$P = \frac{1}{2}\sum_{\nu=0}^{\infty} \log\left[1 + \frac{t^2}{(1+p+\nu)^2}\right], \quad \omega = \Theta(1+p, -t) - t\Psi(1+p),$$

$$\Theta = \sum_{\nu=0}^{\infty}\left[\frac{-t}{1+p+\nu} - \arctan\frac{-t}{1+p+\nu}\right],$$

$$\Psi = \frac{d \log \Gamma(1+p)}{dp} = -\gamma + \sum_{\nu=0}^{\infty} \frac{p}{(\nu+1)(1+\nu+p)},$$

γ = Euler's const.

From these formulas we deduce

$$P(1+p, -t) \to 0 \quad \text{and} \quad \Theta(1+p, -t) \to 0 \quad \text{as} \quad p \to \infty .$$

Dropping the functions P and Q, we obtain

$$C = R\cos\theta, \qquad S = R\sin\theta,$$

$$R = \frac{\Gamma(1+p)}{2\pi} e^{\pi t/2 - p\log(2\pi/a)}, \quad \theta = t\log\frac{2\pi}{a} + \pi\left(\frac{1}{2} + \frac{p}{2} - \frac{2b}{a}\right) - t\Psi(1+p).$$

Replacing $\Gamma(1+p)$ by Stirling's formula we see that $R \to \infty$ as $p \to \infty$. As to the angle θ, the terms containing p may be written

$$p\left[\frac{\pi}{2} - tf(p)\right], \quad f(p) = \sum_{\nu=0}^{\infty} \frac{1}{(\nu+1)(\nu+1+p)}.$$

* See Nielsen, *Handbuch der Gammafunktion*, p. 23.

The positive function $f(p)$ evidently decreases to 0 as a limit as p increases. Hence θ eventually takes the form $\pi p/2+\tau$,

$$\tau = t\left(\gamma + \log\frac{2\pi}{a}\right) + \frac{\pi}{2} - \frac{2\pi b}{a} + \epsilon(p), \quad \epsilon(p) \to 0. \tag{17}$$

We arrive finally at the formulas

$$C = R\cos\left(\frac{\pi p}{2} + \tau\right), \quad S = R\sin\left(\frac{\pi p}{2} + \tau\right), \quad p > p_0, \tag{18}$$

p_0 being a positive number sufficiently large.

These formulas show that, when $k(>0)$ is not too small, the curves (15) cross the line $t=k$ an infinity of times as σ approaches $-\infty$, the functions C and S oscillating between positive and negative values that increase numerically without limit.

The function $Z(s)$ may be replaced by the simpler asymptotic expression

$$Z(s) = Re^{i(\tau-\sigma\pi/2)} \tag{19}$$

in the part of the plane that we are considering.

If t is negative, we derive in the same way formulas exactly similar to (18) and (19), with appropriate changes of sign.

9. **Conjugate functions.** The two functions $Z(a, b, s)$ and $Z(a, a-b, s)$ will be called *conjugate*. Their sum and difference satisfy respectively the simpler functional equations

$$Z_b + Z_{a-b} = x_b(s) = \eta\sum_{n=1}^{\infty}\frac{\cos(2bn\pi/a)}{n^{1-s}}, \tag{20}$$

$$\eta = \frac{2}{\pi}\left(\frac{2\pi}{a}\right)^s \sin\frac{\pi s}{2}\Gamma(1-s),$$

$$Z_b - Z_{a-b} = y_b(s) = \lambda\sum_{n=1}^{\infty}\frac{\sin(2bn\pi/a)}{n^{1-s}}, \tag{21}$$

$$\lambda = \frac{2}{\pi}\left(\frac{2\pi}{a}\right)^s \cos\frac{\pi s}{2}\Gamma(1-s).$$

It is obvious that all functions x_b vanish for $s=-2, -4, \cdots, -2n, \cdots$; and that all functions y_b are integral functions (being regular at $s=1$) and vanish for $s=-1, -3, \cdots, -(2n-1), \cdots$. Hence x_b has the factor $1/\Gamma(s/2+1)$ and y_b has the factor $1/\Gamma((s+1)/2)$.

On account of the simplification of the real zeros of x_b and y_b as compared with those of $Z(s)$, the question arises as to whether there is also a simpli-

fication in the positions of the imaginary zeros, as, for example, if they may lie on a vertical line $\sigma=\sigma_0$. That this is not the case in general may be shown by a single illustration. Namely, the first two imaginary zeros of $Z(5, 1, s) - Z(5, 4, s)$ are found by computation to be at the points $(\sigma, t)=(.63, 8.94)$ and $(.3, 12.15)$.

10. **Linear functions of x_b, or y_b, and their functional equations.** Suppose, now, that a is a fixed, positive integer. We propose to consider the different functions $x_1, x_2, \cdots, x_m$ (or $y_1, \cdots, y_m$), $m=\frac{1}{2}(a-1)$ or $\frac{1}{2}(a-2)$, according as a is odd or even. Let

$$f(s) = a_1x_1 + a_2x_2 + \cdots + a_mx_m \tag{22}$$

be any linear function with constant coefficients. On account of (20), $f(s)$ satisfies the relation

$$f(s) = \eta \sum_{n=1}^{\infty} \frac{a_1 \cos(2\pi n/a) + a_2 \cos(4\pi n/a) + \cdots + a_m \cos(2\pi mn/a)}{n^{1-s}}. \tag{23}$$

The question that we now propose to consider is this: *Can we determine the constants $a_1, a_2, \cdots, a_m$ so that the sum in* (23) *takes the form*

$$k(a_1x_1' + a_2x_2' + \cdots + a_mx_m'), \quad x_\mu' = x_\mu(1-s)?$$

When this is possible the functional equation reduces to the very simple form

$$f(s) = \eta \cdot kf(1-s). \tag{24}$$

Three essentially distinct cases need to be considered:

1. a even, $a=2p$, p odd;
2. a even, $a=4q$;
3. a odd, $a=2m+1$.

11. **Case 1.** Use the abbreviation $c_\mu=\cos(2\pi\mu/a)$. Then the equations of condition for (24) are:

$$\begin{aligned} &c_1a_1 + c_2a_2 + \cdots + c_ma_m = ka_1, \\ &c_2a_1 + c_4a_2 + \cdots + c_{2m}a_m = ka_2, \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ &c_ma_1 + c_{2m}a_2 + \cdots + c_{mm}a_m = ka_m; \end{aligned} \tag{A_1}$$

$$\begin{aligned} &a_1 + a_2 + \cdots + a_m = 0, \\ &a_1 - a_2 + \cdots + (-1)^{m-1}a_m = 0. \end{aligned} \tag{A_2}$$

In the present case $m=p-1$. In order to handle these equations we use the following trigonometric relations which hold for $\mu<p$:

$$c_\mu + c_{2\mu} + \cdots + c_{(p-1)\mu} = \begin{cases} 0, & \mu \text{ odd}, \\ -1, & \mu \text{ even}, \end{cases}$$

$$c_\mu - c_{2\mu} + \cdots - c_{(p-1)\mu} = \begin{cases} 1, & \mu \text{ odd}, \\ 0, & \mu \text{ even}, \end{cases} \tag{25}$$

$$c_\mu^2 + c_{2\mu}^2 + \cdots + c_{(p-1)\mu}^2 = \frac{a}{4} - 1.$$

These are easily verified by first observing that, if $\omega = e^{2\pi i/a}$, any odd power of ω satisfies the equation $1-\omega^\mu+\omega^{2\mu}- \cdots -\omega^{(p-1)\mu}=0$, and any even power satisfies $1+\omega^\mu + \cdots +\omega^{(p-1)\mu}=0$. Again, group the terms thus,

$$(c_\mu \pm c_{p\mu-\mu}) \pm (c_{2\mu} \pm c_{p\mu-2\mu}) + \cdots \text{ (upper, or lower, signs together)}.$$

The sum of the subscripts in each pair is $p\mu$ and hence the two cosines in the same group are equal, if μ is even, or opposite in sign, if μ is odd. The last expression in (25) may be put in the form $\frac{1}{2}[(1-c_{2\mu})+(1-c_{4\mu})+ \cdots]=a/4-1$.

The characteristic equation for the system (A_1) is

$$D(k) \equiv \begin{vmatrix} c_1 - k, & c_2, & \cdots, & c_{p-1} \\ c_2, & c_4 - k, & \cdots, & c_{2p-2} \\ \cdots & \cdots & \cdots & \cdots \\ c_{p-1}, & c_{2p-2}, & \cdots, & c_{(p-1)^2} - k \end{vmatrix} = 0. \tag{26}$$

I shall speak of this relation, and similar relations elsewhere, as the necessary and sufficient condition for a solution of the given system of equations, meaning, thereby, a solution different from $a_1=a_2= \cdots =0$.

We first prove the relation $D(-k)=D(k)$. For this purpose write $D(k)$ in the form

$$\begin{vmatrix} c_{(p-1)(p-1)} - k, & c_{(p-2)(p-1)}, & \cdots, & c_{p-1} \\ c_{(p-1)(p-2)}, & c_{(p-2)(p-2)} - k, & \cdots, & c_{p-2} \\ \cdots & \cdots & \cdots & \cdots \\ c_{p-1}, & c_{p-2}, & \cdots, & c_1 - k \end{vmatrix} \tag{27}$$

obtained by writing rows and columns in the reverse order. The νth element of the μth row satisfies the relation

$$c_{(p-\nu)(p-\mu)} = c_{p(p-\mu-\nu)+\mu\nu} = (-1)^{\mu+\nu-1}c_{\mu\nu}. \tag{28}$$

By means of this, (27) takes the form

$$\begin{vmatrix} -c_1-k, & c_2, & -c_3, & c_4, \cdots, & c_{p-1} \\ c_2, & -c_4-k, & c_6, & -c_8, \cdots, & -c_{p-2} \\ -c_3, & c_6, & -c_9-k, & c_{12}, \cdots, & c_{p-3} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ c_{p-1}, & -c_{2p-2}, & & \cdots, & -c_{(p-1)^2}-k \end{vmatrix}.$$

This reduces to $D(-k)$ by changing the signs of all the elements in the odd numbered columns and the even numbered rows.

Consider now the product determinant $D(k)\,D(-k)$. If $\mu \neq \nu$, the element of the μth row and the νth column is

$$\sum_{b=1}^{p-1} c_{\mu b}c_{\nu b} = \frac{1}{2}\sum_{b=1}^{p-1}[c_{(\mu-\nu)b} + c_{(\mu+\nu)b}] = \frac{1}{2}\Big(\sum\nolimits_1 + \sum\nolimits_2\Big).$$

If $\mu-\nu$ is odd, so also is $\mu+\nu$ and each sum $\sum_1$, or $\sum_2$, is 0, according to (25). If $\mu-\nu$ is even, each sum is -1. If $\mu=\nu$, we obtain an element of the main diagonal of $D(k)D(-k)$, viz.,

$$\sum_{b=1}^{p-1} c_{\nu b}^2 - k^2 = \frac{a}{4} - 1 - k^2 = A.$$

The result is

$$D(k)D(-k) = \begin{vmatrix} A, & 0, & -1, & 0, & -1, \cdots, & -1, & 0 \\ 0, & A, & 0, & -1, & 0, \cdots, & 0, & 1 \\ -1, & 0, & A, & 0, & -1, \cdots, & -1, & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0, & -1, & 0, & -1, & \cdots, & 0, & A \end{vmatrix}.$$

To reduce this determinant, add to the first column the sum of all the other columns. The elements of the first column will then all be equal to $A-\frac{1}{2}(p-3) = \frac{1}{2}-k^2$ and we have $D(k)\,D(-k) = (\frac{1}{2}-k^2)D_1$,

$$D_1 = \begin{vmatrix} 1 & 0 & -1 & 0 & -1 \cdots & -1 & 0 \\ 1 & A & 0 & -1 & 0 \cdots & 0 & -1 \\ 1 & 0 & A & 0 & -1 \cdots & -1 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & -1 & 0 & -1 & 0 \cdots & 0 & A \end{vmatrix}.$$

Now subtract the first row from each of the others and D_1 reduces at once to $(A+1)^{(p-3)/2}\,D_2$,

$$D_2 = \begin{vmatrix} A & -1 & -1 \cdots -1 \\ -1 & A & -1 \cdots -1 \\ \cdot & \cdot & \cdot \quad \cdot \quad \cdot \quad \cdot \\ -1 & -1 & -1 \cdots A \end{vmatrix}.$$

Add to the first column of D_2 the sum of the other columns, take out the common factor $\frac{1}{2}-k^2$, and modify the resulting determinant by adding the first column to each of the others. It then reduces to $(A+1)^{(p-3)/2}$. Collecting the results we have

$$D(k)D(-k) = D^2(k) = \left(\frac{1}{2} - k\right)^2\left(\frac{a}{4} - k^2\right)^{p-3},$$

and hence

$$D(k) = \pm\left(\frac{1}{2} - k^2\right)\left(\frac{a}{4} - k^2\right)^{(p-3)/2}.$$

The solutions $k = \pm 1/2^{1/2}$ must be discarded, since the complete system (A_1) and (A_2) has no solution for these values of k. For, adding the $p-1$ equations (A_1) we obtain

$$-(a_2 + a_4 + \cdots + a_{p-1}) = k(a_1 + a_2 + \cdots + a_{p-1}). \tag{29}$$

Again, changing the signs of every other equation of (A_1) beginning with the second, and adding, we have

$$a_1 + a_3 + a_5 + \cdots + a_{p-2} = k(a_1 - a_2 + a_3 - \cdots - a_{p-1}). \tag{30}$$

Denoting the left members of (29) and (30) by $-\alpha_2$ and α_1 respectively, these equations may be written

$$\begin{aligned} k\alpha_1 + (k+1)\alpha_2 &= 0, \\ (k-1)\alpha_1 + k\alpha_2 &= 0. \end{aligned} \tag{31}$$

The condition for a non-vanishing solution of (31) is $k^2 = \frac{1}{2}$. We have thus separated out the two roots $k = \pm 1/2^{1/2}$ of (26).

Suppose now the system (A), consisting of (A_1) and (A_2), be subjected to the linear transformation

$$\begin{aligned} a_1 + a_3 + \cdots + a_{p-1} &= \alpha_1, \\ a_2 + a_4 + \cdots + a_{p-2} &= \alpha_2, \\ a_\mu &= \alpha_\mu, \quad \mu = 3,4, \cdots, p-1. \end{aligned}$$

Suppose, further, that the system (A_1) in the new unknowns $\alpha_1, \alpha_2, \cdots$ be replaced by the two equations (31) together with $p-3$ of the equations

(A_1) so selected that the new system (A_1') of $p-1$ equations in $\alpha_1, \cdots, \alpha_{p-1}$ are linearly independent. The two equations (A_2) may be replaced by $\alpha_1=0$, $\alpha_2=0$. Consequently, if we drop the terms containing α_1 and α_2 in the new system (A_1'), it is obvious that the $p-3$ equations other than (31) will have as characteristic determinant one of degree $p-3$ in k whose roots will be $k=\pm\frac{1}{2}a^{1/2}$ since the characteristic equation is unaltered by linear transformation. These $p-3$ equations of (A_1') will consequently have no solution, if $k=\pm 1/2^{1/2}$, and hence the complete system (A) has no solution for these values of k.

Since each of the roots $\pm\frac{1}{2}a^{1/2}$ of (26) is of multiplicity $\frac{1}{2}(p-3)$, it follows that for either of these values of k the system (A_1) has only $p-1-\frac{1}{2}(p-3)-2 = \frac{1}{2}(p-3)$ linearly independent equations, since there are two relations (29) and (30) that are satisfied by (A_2). We accordingly obtain the following result:

THEOREM 3. *There are $\frac{1}{2}(p-3)$ linearly independent functions $f(s)$ that satisfy the functional equation*

$$f_1(s) = a^{1/2}\left(\frac{2\pi}{a}\right)^s \sin\frac{\pi s}{2}\,\frac{\Gamma(1-s)}{\pi} f_1(1-s), \tag{I}$$

and likewise $\frac{1}{2}(p-3)$ linearly independent functions that satisfy the equation

$$f_2(s) = -a^{1/2}\left(\frac{2\pi}{a}\right)^s \sin\frac{\pi s}{2}\,\frac{\Gamma(1-s)}{\pi} f_2(1-s). \tag{II}$$

12. **Case 1, the y-relations.** We next consider the function

$$f(s) = b_1y_1 + b_2y_2 + \cdots + b_my_m \tag{32}$$

and seek to determine $b_1, \cdots, b_m$ so that formula (21) reduces to

$$f(s) = \lambda \cdot kf(1-s). \tag{33}$$

Using for brevity $s_\mu = \sin(2\mu\pi/a)$, the equations of condition are

$$\begin{aligned} s_1b_1 + s_2b_2 + \cdots + s_mb_m &= kb_1, \\ s_2b_1 + s_4b_2 + \cdots + s_{2m}b_m &= kb_2, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots & \\ s_mb_1 + s_{2m}b_2 + \cdots + s_{mm}b_m &= kb_m. \end{aligned} \tag{B}$$

Denoting the determinant of this system by $D(k)$, we prove, just as in §11, $D(-k)=D(k)$, since $s_{(p-\nu)(p-\mu)}$ satisfies (28). Using the relations

$$\sum_{b=1}^{m} s_{\mu b} s_{\nu b} = \tfrac{1}{2} \sum_{b=1}^{m} [c_{(\nu-\mu)b} - c_{(\nu+\mu)b}] = 0, \quad \mu \neq \nu,$$

$$\sum_{b=1}^{m} s_{\nu b}^2 = \tfrac{1}{2} \sum (1 - c_{2\nu b}) = \frac{a}{4}, \tag{34}$$

the product $D(k)D(-k) = D^2(k)$ reduces at once to $(a/4 - k^2)^{p-1}$. From this we deduce the following result:

THEOREM 4. *There are* $\frac{1}{2}(p-1)$ *linearly independent functions* (32) *satisfying the relation*

$$f_3(s) = a^{1/2} \left(\frac{2\pi}{a}\right)^s \cos \frac{\pi s}{2} \frac{\Gamma(1-s)}{\pi} f_3(1-s), \tag{III}$$

and $\frac{1}{2}(p-1)$ *linearly independent functions that satisfy*

$$f_4(s) = -a^{1/2} \left(\frac{2\pi}{a}\right)^s \cos \frac{\pi s}{2} \frac{\Gamma(1-s)}{\pi} f_4(1-s). \tag{IV}$$

13. **Case 2,** $a = 4q$. Consider first the function (32) and the conditions (B) with $m = 2q-1$. We find, just as in §12, $D(k)D(-k) = (a/4 - k^2)^{2q-1}$, by using (34) which holds in this case on account of (35). But we no longer have the relation $D(-k) = D(k)$ to determine the multiplicity of each root. The fact appears to be that $\frac{1}{2}a^{1/2}$ is a root of $D(k)$ of multiplicity q, and $-\frac{1}{2}a^{1/2}$ is of multiplicity $q-1$. I have not succeeded in obtaining a complete proof for this statement, but the following considerations will indicate the reasons for it.

Denote by $(B_1), (B_2), \cdots, (B_m)$ the separate equations of (B) in the order in which they are written. Multiply these by $s_\mu, s_{2\mu}, \cdots, s_{m\mu}$ respectively and add. We obtain $\frac{1}{4}ab_\mu = k(s_\mu b_1 + \cdots + s_{m\mu} b_m)$, which, by using (B_μ) in the right member, reduces to

$$\left(\frac{a}{4} - k^2\right) b_\mu = 0 \qquad (\mu = 1, 2, \cdots, q-1). \tag{B_μ'}$$

It further seems to be always possible to form, by additions or subtractions, a linear combination of the odd numbered equations (B_μ) in case $q \equiv 1$ or 2 (mod 4), or of the even numbered equations in case $q \equiv 0$ or 3 (mod 4), which reduce to the form

$$\begin{aligned} &(\epsilon_1 s_1 + \epsilon_3 s_3 + \cdots - k)(\epsilon_1 b_1 + \epsilon_3 b_3 + \cdots) = 0, \\ &(\epsilon_2 s_2 + \epsilon_4 s_4 + \cdots - k)(\epsilon_2 b_2 + \epsilon_4 b_4 + \cdots) = 0, \end{aligned} \tag{B_q' or}$$

the ϵ's having the values 1, -1, or 0 ($\epsilon_q = \pm 2$, or 0), such that the first

factor reduces by Gauss' sums (or otherwise) to $(\frac{1}{2}a^{1/2}-k)$. The system of $2q-1$ equations (B_μ), (B'_μ), (B'_q), $\mu=1, \cdots, q-1$, appears to be equivalent to the original system (B) and evidently has a determinant whose factors in k are $(a/4-k^2)^{q-1}$ $(a^{1/2}/2-k)$.

To each of the two values of k correspond functions f(s) having as many independent parameters b_μ as the multiplicity of the corresponding root k and satisfying one or the other of the relations (III), (IV).

Consider now the functions (23) and the system (A) that their coefficients must satisfy. In evaluating the product $D(k)D(-k)$, in this case, we use the relations

$$
\begin{aligned}
&c_\mu + c_{2\mu} + \cdots + c_{(2q-1)\mu} = \begin{cases} 0, & \mu \text{ odd}, \\ -1, & \mu \text{ even}, \end{cases} \\
&c_\mu - c_{2\mu} + \cdots + c_{(2q-1)\mu} = \begin{cases} 0, & \mu \text{ odd}, \\ 1, & \mu \text{ even}, \end{cases} \\
&c_\mu^2 + c_{2\mu}^2 + \cdots + c_{(2q-1)\mu}^2 = \frac{a}{4} - 1 = A, \\
&\sum_{b=1}^{2q-1} c_{\mu b} c_{\nu b} = \begin{cases} 0, & \mu - \nu \text{ odd}, \\ -1, & \mu - \nu \text{ even}. \end{cases}
\end{aligned} \tag{35}
$$

The product $D(k)D(-k)$ is of the same form as in §11 except that, the determinant being of odd order, the elements 0, -1 in the last row and column will interchange.

To reduce this determinant, add all the remaining odd numbered columns to the first column, and all the other even numbered columns to the second one. We can then divide $k^2(k^2-1)$ out of the first two columns. The resulting determinant at once reduces, after adding the first column to each of the other odd numbered columns and the second column to each of the other even numbered columns, and we obtain

$$D(k)D(-k) = k^2(k^2-1)(\tfrac{1}{4}a - k^2)^{2q-3}.$$

Continuing with the methods of Case 1, it is easy to show that $D(k)$ has the roots 0 and -1, which must be excluded. The remaining $2q-3$ roots, apparently of multiplicity $q-1$ for $\frac{1}{2}a^{1/2}$ and $q-2$ for $-\frac{1}{2}a^{1/2}$, determine functions $f(s)$ having as many independent parameters a_μ as the multiplicity of the corresponding root k and satisfying (I), or (II). The facts that indicate the degree of multiplicity are similar to those given at the beginning of this article.

14. **Case 3, $a=2m+1$.** Considering first the conditions (B) that the function (32) shall satisfy (33), we readily find $D(k)D(-k)=(a/4-k^2)^m$. The same considerations as in the preceding case lead to the belief that $D(k)$ has the two roots $\pm\frac{1}{2}a^{1/2}$ with the same degree of multiplicity, if m is even, and the root $\frac{1}{2}a^{1/2}$ with multiplicity one higher, in case m is odd.*

When m is odd, the relation corresponding to (B_q') of §13 is more easily found as it consists of a sum of the equations (B_μ) multiplied by $\epsilon_\mu=1, -1$, or 0, $\mu=1, \cdots, m$.

The conditions that the function (22) shall satisfy (24) are the same as for Case 1, except that (A_2) has only one equation $a_1+a_2+\cdots=0$. The trigonometric relations in this case are

$$c_\mu + c_{2\mu} + \cdots + c_{m\mu} = -\frac{1}{2},$$
$$c_\mu^2 + c_{2\mu}^2 + \cdots + c_{m\mu}^2 = \frac{a}{4} - \frac{1}{2}.$$

The product $D(k)D(-k)$ consists of $a/4-\frac{1}{2}-k^2$ for each element in the main diagonal and $-\frac{1}{2}$ for every other element. By adding all the other rows to the first, the common element $\frac{1}{4}-k^2$ divides out. By adding one-half the elements of the first row of the resulting determinant to each of the other rows we obtain

$$D(k)D(-k) = \left(\frac{1}{4} - k^2\right)\left(\frac{a}{4} - k^2\right)^{m-1}.$$

By adding all the equations (A_1) we get $(-\frac{1}{2}-k)(a_1+a_2+\cdots+a_m)=0$, which shows that $D(k)$ has the root $-\frac{1}{2}$. This root does not satisfy the complete system (A).

Apparently $k=\frac{1}{2}a^{1/2}$ is a root of $D(k)$ of multiplicity one greater than that of the root $-\frac{1}{2}a^{1/2}$, if m is even, and of the same multiplicity if m is odd. The results of Cases 2 and 3 may be thus summarized:

The number of linearly independent functions satisfying (I) *and* (III), *or* (II) *and* (IV), *is equal to the multiplicity of the root* $\frac{1}{2}a^{1/2}$ *or* $-\frac{1}{2}a^{1/2}$, *of the determinant of the corresponding system of equations* (A_1) *and* (B).

15. **The argument of $f_\mu(s)$ on the line $\sigma=\frac{1}{2}$.** The functional equation (I) may be put in the form

* Case 3, for a a prime number, has been completely solved by E. Cahen, *Sur la fonction $\zeta(s)$ de Riemann et sur des fonctions analogues*, Annales de l'Ecole Normale Supérieure, 1894, pp. 75–164 His method, which is particularly suitable when a is a prime, does not seem to be applicable to any other case. We might notice here some errors in Cahen's memoir. On p. 151, last line, the coefficients of the functions that we have denoted by y_1 and y_2 should be interchanged. A similar interchange should be made in line 4, p. 152.

$$\frac{f_1(1-s)}{f_1(s)} = \frac{2}{a^{1/2}}\left(\frac{a}{2\pi}\right)^s \cos\frac{\pi s}{2}\,\Gamma(s). \tag{36}$$

Let $s=\frac{1}{2}+it$ and write $f_1(\frac{1}{2}+it)=\rho_1(t)e^{i\phi_1(t)}$. If we take the logarithm of (36), the left member becomes $-2\phi_1(t)$. By using the procedure of Gram* as developed for $\zeta(s)$, we obtain

$$\frac{-\phi_1(t)}{\pi} = \frac{t}{2\pi}\left[\log\left(\frac{at}{2\pi}\right)-1\right]-\frac{1}{8}. \tag{37}$$

The functional equations (II), (III), (IV) yield similar formulas in which ϕ_1 is replaced by ϕ_2, ϕ_3, ϕ_4 respectively, and the term $-\frac{1}{8}$ is replaced by $\frac{3}{8}$, $\frac{1}{8}$, $\frac{5}{8}$ respectively.

Let $f(s)$, without subscript, denote any one of the functions $f_\mu(s)$. Write $f(\frac{1}{2}+it)$ in the form

$$f(\tfrac{1}{2}+it) = \rho\cos\phi + i\rho\sin\phi = C(t)+iS(t).$$

Denote by γ any value of t for which $\sin\phi(t)$ is 0, and by β any value of t for which $\cos\phi(t)$ is 0. If there are any real values $t=\alpha\neq\gamma$ which make $S(t)$ vanish, they must be zeros of $\rho(t)$ and hence zeros of $f(\frac{1}{2}+it)$ as well as of $C(t)$. The result obtained by Gram for $\zeta(\frac{1}{2}+it)$ holds equally well for our functions $f(\frac{1}{2}+it)$, viz.:

THEOREM 5. *If $\gamma_n(>2\pi/a)$ and γ_{n+1} are two consecutive roots of* $\sin\phi(t)$, *and if $C(\gamma_n)$ and $C(\gamma_{n+1})$ have the same sign, then an odd number of roots α of $f(\frac{1}{2}+it)$ occur between γ_n and γ_{n+1} on the line $\sigma=\frac{1}{2}$, and hence there is at least one root in this interval. Similarly, if $S(\beta_n)$ and $S(\beta_{n+1})$ have the same sign, there is at least one root α in the interval (β_n,β_{n+1}).*

A simpler proof of Theorem 5 than that given by Gram would be this. Since γ_n and γ_{n+1} are two consecutive roots of $\sin\phi(t)=0$ and since $-\phi(t)$ is an increasing function for $at>2\pi$, then $\phi(\gamma_n)$ and $\phi(\gamma_{n+1})$ are two consecutive multiples of π and hence $\cos\phi(t)$ takes opposite signs at $t=\gamma_n$ and γ_{n+1}. Hence, if $C(t)=\rho(t)\cos\phi(t)$ preserves the same sign for two consecutive values of γ, the factor $\rho(t)$ must change sign in the interval (γ_n,γ_{n+1}).

16. **The number of imaginary zeros of $f(s)$.** The functional equations (I), $\cdots$, (IV) may be written

(I′) $\chi_1(s)=\chi_1(1-s)$,

$$\chi_\mu(s) = f_\mu(s)\Gamma\left(\frac{s}{2}\right)\left(\frac{a}{\pi}\right)^{s/2}, \quad \mu=1,2;$$

(II′) $\chi_2(s) = -\chi_2(1-s)$,

* J. P. Gram, loc. cit., pp. 298–304.

(III′) $\chi_3(s) = \chi_3(1-s)$,

$$\chi_\mu(s) = f_\mu(s)\Gamma\left(\frac{1+s}{2}\right)\left(\frac{a}{\pi}\right)^{s/2}, \quad \mu = 3,4.$$

(IV′) $\chi_4(s) = -\chi_4(1-s)$,

We propose to determine an upper limit $N(T)$ for the number of imaginary zeros of the functions $f_\mu(s)$, whose coefficients we assume to be all real, in the region $0<t<T$.

Consider first the function

$$\xi_\mu(s) = \tfrac{1}{2}s(s-1)\chi_\mu(s), \quad \mu = 1,2.$$

The zeros of $\xi_\mu(s)$ (excepting $s=\frac{1}{2}$) are the imaginary zeros of $f_\mu(s)$, $\mu=1, 2$. They are symmetrical with respect to the lines $t=0$ and $\sigma=\frac{1}{2}$. Let $f(s)$ denote any one of the functions under consideration and write

$$f(s) = \sum_{n=n_0}^{\infty} c_n n^{-s},$$

the c_n all real. Let σ_0 be the least positive value of σ such that

$$|c_0 n_0^{-s}| > \sum_{n=n_0+1}^{\infty} |c_n n^{-s}|.$$

Then there are no imaginary zeros of $f(s)$ in the half-plane $\sigma>\sigma_0$ and, on account of the functional equations, none in the half-plane $\sigma<1-\sigma_0$.

The method of evaluating $N(T)$ is very similar to that employed for $\zeta(s)$. I follow as closely as possible the treatment of Backlund* and refer to his paper for details.

Consider the rectangle whose vertices are $\alpha \pm iT$, $1-\alpha \pm iT$, α real $>\sigma_0$. Then we have

$$\pi N(T) = \Delta_{\alpha\beta\gamma} \arg \xi_\mu(s),$$

the right member denoting the increment that the argument of $\xi_\mu(s)$ acquires as s describes the broken line $\alpha\beta\gamma$, $\beta=\alpha+iT$, $\gamma=\frac{1}{2}+iT$.

The only variations from the steps given by Backlund arise in the calculation of $\Delta_{\alpha\beta\gamma} \arg f(s)$. Denoting by $C(\sigma, t)$ the real part of $f(s)$, then $\Delta_{\alpha\beta\gamma} \arg f(s)/\pi$ will not exceed the number of times that $C(\sigma, t)$ vanishes on the broken line $\alpha\beta\gamma$. The results of §7 tell us that, if $f(s)$ contains the term

* R. J. Backlund, *Ueber die Nullstellen der Riemannschen Zetafunktion*, Acta Mathematica vol. 41 (1918), p. 348.

$Z(a, 1, s)$, the curve $C(\sigma, t)=0$ has no right-hand asymptotes and the vertical line $\sigma=\alpha>\sigma_0$ does not meet the curve. In this case $\Delta_{\alpha\beta} \arg f(s)<\pi/2$.

In other cases the lines $2t \log n_0=(2n+1)\pi$ are right-hand asymptotes of the curve $C=0$ and hence $C(\sigma, t)$ vanishes on the line $\sigma=\alpha$ at the points

$$t=\frac{(2n+1)\pi}{2\log n_0}+\epsilon(\alpha,n), \quad \lim_{\alpha\to\infty}\epsilon(\alpha,n)=0. \tag{38}$$

Moreover, when α is sufficiently large, the change of sign of $C(\alpha, t)$, as t increases through one of the values (38), is due solely to the first term $c_{n_0}[\cos (t \log n_0)]n_0^{-\alpha}$ and this term obviously changes sign but once in the vicinity of each point (38).

Let n_1 be the largest value of n in (38) for which $t<T$. Then we obtain

$$n_1<\frac{T\log n_0}{\pi}+b_1$$

and hence

$$\Delta_{\alpha\beta}\arg f(s)<T\log n_0+b_2,$$

b_1 and b_2 being constants.

We obtain as the final result,

$$N(T)=\frac{T}{2\pi}\left[\log\frac{an_0^2T}{2\pi}-1+\epsilon(T)\right], \quad \lim_{T\to\infty}\epsilon(T)=0. \tag{39}$$

The functions $\chi_\mu(s)$, $\mu=3, 4$, are regular at $s=0, 1$. Treating them in the same way as $\xi_1(s)$, $\xi_2(s)$, we arrive at (39) as before.

17. **Proof that the functions $\chi_\mu(s)$ have an infinity of roots.** Denoting by ρ_n the roots of any of the functions $\chi_\mu(s)$ which are situated in the half-plane $t>0$, we find from (39) (Backlund, p. 354) that $\sum|\rho_n^{-1-\epsilon}|$ is convergent, ϵ being a fixed positive number as small as we please. Hence the exponent of convergence* of the roots of $\chi_\mu(s)$ does not exceed 1.

If we substitute $s=\frac{1}{2}+it$ in the functions $\xi_\mu(s)$, $\chi_\mu(s)$, the functional equations (I'), $\cdots$, (IV') show that the functions ξ_1, ξ_2/t, χ_3, χ_4/t are uniform integral functions of $t^2=z$. The argument of Hadamard† relative to $\zeta(s)$ applies at once to these functions and shows that each of them is of genus zero as a function of z and hence has an infinity of zeros, since it is obviously not a polynomial.

Moreover, we can prove the following:

---

* E. Borel, *Leçons sur les Fonctions Entières*, p. 18.

† Borel, loc. cit., 2d edition, pp. 84–88.

THEOREM 6. *Those functions $f_\mu(s)$ that contain the term $Z(a, 1, s)$ have an infinity of roots on the line $\sigma = \frac{1}{2}$.*

The proof will be omitted as it is very similar to the Hardy and Littlewood* proof for $\zeta(s)$. We have followed the details as given by E. Landau.† They apply with no change for the functions $f_\mu(s)$, $\mu = 1, 2$, except in the values of some of the multiplying constants, and in using for $f_\mu(s)$ a series of the form

$$\sum_{n=1}^{\infty} c_n n^{-s}, \quad c_1 \neq 0, \tag{40}$$

absolutely convergent for $\sigma = 1$, in place of the series used for $\zeta(s)$.

As to the functions $f_\mu(s)$, $\mu = 3, 4$, which we also represent by (40), on account of the presence of the factor $\Gamma\{\frac{1}{2}(1+\beta)\}$ in the corresponding functional equations (III′) and (IV′), all the formulas depending on this factor must be subjected to a re-evaluation, the general effect of which is to increase the exponent of t by $\frac{1}{2}$ in all these formulas. This extra power of t divides out when we come to solve formula (526) of Landau, giving the same result $O(t^{7/8})$ as for $\zeta(s)$.

It is to be remarked that $\xi_2(\frac{1}{2}+it)$ and $\chi_4(\frac{1}{2}+it)$ have the root $t=0$ and accordingly the functions $f_2(s)$ and $f_4(s)$ have the real root $s=\frac{1}{2}$ in addition to the infinity of real roots already determined in §9.

18. **Generalization of the functional equation.** Instead of the forms assumed for (24) and (33) we could make a more general assumption in which $f(1-s)$ in the right member is replaced by $\bar{f}(1-s)$, the notation $\bar{f}$ meaning that the coefficients $a_1, \cdots, a_m$ (or $b_1, \cdots, b_m$) in f are replaced by their conjugate imaginary values $\bar{a}_1, \cdots$.

Instead of the equations of condition (A) or (B), we would have similar equations $(\bar{\text{A}})$ or $(\bar{\text{B}})$ in which the letters $a_1, \cdots$ in the right members are replaced by $\bar{a}_1, \cdots$. Each case could be treated as in §13. For example, half the $(\bar{\text{B}})$ equations of Case 1 could be replaced by the equations obtained by multiplying the given equations by $s_\mu, s_{2\mu}, \cdots$, respectively, adding, and using the conjugate of the μth equation $(\bar{\text{B}})$. This would give

$$(a/4 - k\bar{k})b_\mu = 0 \qquad (\mu = 1, 2, \cdots, (p-1)/2).$$

From this we see that the system $(\bar{\text{B}})$ is satisfied if $k = \frac{1}{2}a^{1/2}e^{i\theta}$, θ arbitrary.

---

* *The Riemann zeta-function and the distribution of primes*, Acta Mathematica, vol. 41 (1917), pp. 177–184.

† *Vorlesungen über Zahlentheorie*, vol. 2, 1927, pp. 78–85.

It is obvious that the resulting functional equation, multiplied by its conjugate, will give the relation

$$F(s) = \lambda^2 \cdot \frac{a}{4} F(1-s), \quad F(s) = f(s)\bar{f}(s).$$

Functions $f(s)$ satisfying equations obtained in this way include as special cases the L-functions with imaginary characters.

CORNELL UNIVERSITY,
ITHACA, N. Y.

CONCERNING ZERO-DIMENSIONAL SETS IN EUCLIDEAN SPACE*

BY

R. L. WILDER

A point set M, lying in a euclidean space, E_n, of n dimensions, is said to be zero-dimensional in the Menger-Urysohn sense, if for every point P of M and every positive number ϵ there exists a separation of M into two mutually separated sets M_1 and M_2 such that M_1 contains P and the diameter of M_1 is less than ϵ.†

The present paper is intended to serve as a contribution to the study of zero-dimensional sets in E_n, particularly with reference to the relations of these sets to their complements. The property of accessibility of a point set from all sides is introduced and it is shown that in E_n $(n>1)$ all zero-dimensional sets possess this property and in E_2 are characterized by it. An example is given to show that, in the definition of accessibility used, arcs cannot be employed instead of continua. The same example shows that if M is a zero-dimensional set in E_2, then, although it is well known that M is homeomorphic with a subset of I_2, the set of all points in E_2 both of whose coördinates are irrational, there does not in general exist a one-to-one continuous transformation of E_2 into itself which carries M into a subset of I_2; i.e., M is not in general isotopic with a subset of I_2. A necessary and sufficient condition is then obtained under which M will be isotopic with a subset of I_2, and by application of this condition it is found that every punctiform F_σ in E_2 is isotopic with a subset of I_2.

1

DEFINITION. In this paper, the term *region* will be used to denote a simply connected, bounded domain.

It follows from a theorem of Sierpinski‡ that zero-dimensional sets are

---

* Presented to the Society, April 15, 1927, under the title *Concerning zero-dimensional sets in the plane*; received by the editors March 31, 1928.

† For a general summary of the Menger-Urysohn dimension theory and references, see K. Menger, *Bericht über die Dimensionstheorie*, Jahresbericht der Deutschen Mathematiker-Vereinigung, vol. 35 (1926), pp. 113–150.

‡ W. Sierpinski, *Sur les ensembles connexes et non connexes*, Fundamenta Mathematicae, vol. 2 (1921), pp. 81–95.

identical with those sets that are punctiform and homeomorphic* with linear sets. Mazurkiewicz has shown† that if a_1 and a_2 are points of a punctiform set A in E_n which is homeomorphic with a linear set, and D is a domain of E_n containing a_1, then there exists a continuum C which lies wholly in D, contains no point of A and separates a_1 from a_2. Hence, if M is a zero-dimensional set in E_2, P and Q are distinct points of M, ϵ is any positive number less than the distance from P to Q, and D is the set of all points whose distance from P is less than ϵ, there exists a set F which is a subcontinuum of $D-D\times M$ and which separates P from Q. If R is that component‡ of E_2-F determined by P, then R is a domain. Its boundary, B, is a continuum by virtue of a theorem of Brouwer;§ hence R is simply connected‖ and consequently a region as defined above, its boundedness being evident. That B contains no point of M is obvious, since B is a subset of F.

If a point set M in E_2 has the property that for every point P of M and every positive number ϵ there exists a region containing P every point of which is at a distance from P less than ϵ and whose boundary contains no point of M, then it is obvious that M is zero-dimensional. We have then the following lemma:

LEMMA 1. *In order that a point set M in E_2 should be zero-dimensional it is necessary and sufficient that for every point P of M and every positive number ϵ there exist a region containing P every point of which is at a distance from P less than ϵ and whose boundary contains no point of M.*

DEFINITION. If P is a point of a point set M in E_2, then M will be said to be *locally separated* at P provided that for every positive number ϵ there exists a region containing P whose diameter is less than ϵ and whose boundary contains no point of M. If M is locally separated at all of its points, then M will be called locally separated.

---

* Two sets M and N are said to be *homeomorphic* if there exists a one-to-one continuous correspondence between them. A set is called *punctiform* if it contains no continuum. (A continuum is a closed and connected point set containing more than one point.)

† S. Mazurkiewicz, *Sur un ensemble G_δ, punctiforme, qui n'est pas homéomorphe avec aucun ensemble linéaire*, Fundamenta Mathematicae, vol. 1 (1920), pp. 61–81, Theorem IV. It is clearly intended that the word "punctiforme" appear in this theorem—thus: "A est un ensemble *punctiforme* de $R_0 \cdots$."

‡ If M is a point set and P a point of M, then that component of M determined by P is the set of all points $\{x\}$, of M, such that x and P lie in a connected subset of M.

§ Cf. L. E. J. Brouwer, *Beweis des Jordanschen Kurvensatzes*, Mathematische Annalen, vol. 69 (1910), pp. 169–175, Theorem 3.

‖ Cf. R. L. Moore, *Concerning continuous curves in the plane*, Mathematische Zeitschrift, vol. 15 (1922), pp. 254–260, Theorem 2.

DEFINITION. A sequence of regions, G, is said to *close down* on a point P if every region of G contains P and if for every positive number ϵ all but a finite number of regions of G are of diameter less than ϵ.

DEFINITION. A set of regions, G, is said to *cover* a point set M *in the Vitali sense* provided that if P is any point of M there exists an infinite sequence of regions of G closing down on P.

DEFINITION. A set of regions, G, is said to have *property* H if for every positive number ϵ there exist only a finite number of regions of G of diameter greater than ϵ.

DEFINITION. A point P of a set M in E_2 will be said to be *accessible from all sides* provided that if D is a Jordan region* whose boundary contains P, then P can be joined to any point x of D by a continuum C which lies wholly in D except for P, and such that all points of C except P and possibly x are points of E_2-M. If in this definition the words "a continuum" are replaced by "an arc," the point P will be called *arcwise accessible from all sides*.

DEFINITION. If M is a set of points in E_2, then E_2 will be said to be *accessible from all sides with respect to* M if every point P of E_2 is accessible from all sides when P is added to M. An analogous definition for "arcwise accessible from all sides with respect to M" is obvious.

DEFINITION. Two sets M and N lying in spaces S and T, respectively, are said to be *isotopic* in case there exists a one-to-one continuous correspondence between S and T under which M and N correspond to one another. Obviously S and T can be the same space.

DEFINITION. If P is a point of a point set M in E_2, then M will be said to be *simply locally separated* at P provided that for every positive number ϵ there exists a region containing P whose diameter is less than ϵ and whose boundary is a simple closed curve which contains no point of M. If M is simply locally separated at all of its points, then M will itself be called simply locally separated.

LEMMA 2. *If M is a zero-dimensional set, and P is a point not belonging to M, then $M+P$ is zero-dimensional.*

This lemma is an immediate consequence of a result due to Urysohn,† to the effect that the set of points at which a set T is of dimension $\geqq 0$ is dense in itself.

---

* A Jordan region is a bounded domain complementary to a simple closed curve.

† P. Urysohn, *Sur les multiplicités Cantoriennes*, Fundamenta Mathematicae, vol. 7 (1925), pp. 30–137, and vol. 8 (1926), pp. 225–359. See vol. 8, pp. 272–273.

THEOREM 1. *If M is a zero-dimensional set in E_2 there exists a sequence of regions G whose boundaries contain no points of M, which cover M in the Vitali sense, and such that G has property* H.*

Consider first a bounded zero-dimensional set M. Let M' denote the set composed of M together with all its limit points. By Lemmas 1 and 2, if P is any point of M', there exists a sequence of regions, $G(P)$, closing down on P whose boundaries contain no point of M. Let G' be the collection of all regions belonging to sequences of the type $G(P)$. By the Borel Theorem there exists a finite set, G_1, of regions of G' covering M'. From G' omit all regions of diameter >1 and call the resulting set of regions G_1'. As G_1' covers M', there exists a finite subset, G_2, of G_1', which covers M'. From G_1' omit all regions of diameter $>\frac{1}{2}$ and call the resulting set of regions G_2'. In general, if G_n' consists of the set of all regions of G' of diameter $\leqq 1/n$, there exists a finite subset, G_{n+1}, of G_n', which covers M'.

Let $G=\sum_1^\infty G_n$. Then if ϵ is any positive number, only a finite number of regions of G are of diameter $>\epsilon$, and furthermore, since for every point P of M there exists a region of G_n $(n=1, 2, 3, \cdots)$ covering P, G covers M in the Vitali sense.

Since every unbounded set in E_2 is the sum of a denumerable collection of bounded sets, the set M, if unbounded, is the sum of a sequence of bounded sets $M_1, M_2, M_3, \cdots$. The set G' can be selected as before; in general G_n' can consist of all regions of G' of diameter $\leqq 1/n$ and G_{n+1} can be a finite set of regions of G_n' covering the set $\sum_1^n M_i$ together with its limit points.

THEOREM 1a. *If M is a simply locally separated set in E_2, there exists a set of Jordan regions G whose boundaries contain no point of M, which cover M in the Vitali sense, and such that G has property* H.

THEOREM 2. *If M is a zero-dimensional set in E_2, then E_2 is accessible from all sides with respect to M.*

Let D be a Jordan region and P a point on the boundary, B, of D. Let Q be any point of D. There exists a simple closed curve J which contains P and Q and which lies, except for P, wholly in D. The curve J is the sum of two arcs, t_1 and t_2, which have in common only the points P and Q.

Let z be a point of R, the region bounded by J, and let t be an arc whose end points are P and Q, which contains z, and which lies, except for P and Q, wholly in R. Let M_1 be the set of points common to M and $J-(P+Q)$.

* Note added in proof-reading: Since I have recently shown (in my paper *Concerning the Phragmen-Brouwer Theorem*, presented to the American Mathematical Society, December 27, 1928) that the theorem of Brouwer referred to above extends to n dimensions $(n>2)$, it is clear that Lemma 1 and consequently Theorem 1 are true in E_n, for $n>2$.

As M is zero-dimensional, there exists, by Theorem 1, a set of regions G which covers M in the Vitali sense and has property H, and such that the boundaries of regions of G contain no points of M.

If x is any point of M_1, there exists a region of G, $g(x)$, which contains x and such that $g(x)$ together with its boundary lies wholly in D and contains no point of t. The set of all points contained in regions of the type $g(x)$, together with their boundaries, denote by T. That the set of points $T+J$ is a continuum C is easily shown from the properties of G.

Let that connected domain complementary to C, which contains z, be denoted by $D(z)$. The boundary, F, of $D(z)$ is a subset of C and that it contains P and Q is easily seen from the fact that all of t, except P and Q, lies in $D(z)$. That F is a continuum follows from the Brouwer theorem referred to above, and that it contains no point of M except possibly P and Q is obvious. Hence F is a continuum which contains P and Q, contains no point of M except possibly P and Q, and lies, except for P, wholly in D.

THEOREM 3. *If M is a zero-dimensional set in E_2 and P and Q are any two points of E_2, then P and Q can be joined by a continuum K every point of which, except possibly P and Q, is in E_2-M; and indeed, if J is any simple closed curve enclosing both P and Q, K may be selected so as to lie entirely within J.*

From the above it is evident that whereas, by a theorem of Sierpinski,* the complement of a punctiform set in E_2 is connected im kleinen,† the complement of a set having the stronger property of zero-dimensionality is strongly connected im kleinen.

If, when considering space $E_n(n>2)$, we define accessibility from all sides as above, except that Jordan regions are replaced by bounded domains complementary to n-dimensional spheres, the above results are easily extended to higher spaces. Thus, we have the following theorem:

THEOREM 4. *If M is a zero-dimensional set in $E_n(n>1)$ then all points of E_n are accessible from all sides with respect to M, and the complement, E_n-M, is strongly connected im kleinen.*

I shall merely indicate how the proof is given for E_3. Let S be a sphere, P

---

* W. Sierpinski, *Sur un ensemble punctiforme connexe*, Fundamenta Mathematicae, vol. 1 (1920), pp. 7–10.

† A set M is called connected im kleinen provided that if P is any point of M and ϵ is any positive number, there exists a positive number ρ such that if Q is a point at a distance from P less than ρ, there exists a connected subset N of M containing both P and Q every point of which is at a distance from P less than ϵ. If N can always be taken to be a continuum, then M is called strongly connected im kleinen.

a point on S and Q a point within S. Let T be any plane passing through P and Q. The intersection of S with T is a circle C, and with M is a point set m. The set m is zero-dimensional, and by Theorem 3 there exists a continuum K lying entirely within C, on T, and containing P and Q but no points of m except possibly P and Q. The rest of the proof should be obvious.

THEOREM 4a. *If, in $E_n(n>1)$, D is a connected domain and M is a zero-dimensional set and P and Q are distinct points of D, then there exists, in D, a continuum C which contains P and Q but which contains no point of M, except possibly P and Q.*

I shall indicate the proof for E_3. (For E_2 use Theorem 3 and the notion of simple chain indicated below.) Every point x of D is the center of a sphere S_x which lies wholly in D. Let G denote the collection of all such spheres. Then there exists, from P to Q, a simple chain, $S_1, S_2, \cdots, S_k$, of spheres* of the collection G. For each $i(i=1, 2, \cdots, k-1)$ let P_i denote a point of E_3-M common to S_i and S_{i+1}. By passing a plane through P_i and P_{i+1} and proceeding as in the proof of Theorem 4, it can be shown that there exists a continuum C_{i+1} which lies wholly in S_{i+1}, contains P_i and P_{i+1}, but no point of M. Similar continua C_1 and C_k can be obtained, where C_1 joins P and P_1 in S_1, and C_k joins P_{k-1} and Q in S_k. The continuum $C=\sum_1^k C_i$ fulfills the condition stated in the theorem.†

COROLLARY. *In $E_n(n>1)$ the complement of a zero-dimensional set is strongly connected.*

THEOREM 5. *In order that a set in E_2 should be zero-dimensional it is necessary and sufficient that it should be accessible from all sides.*

That the condition is necessary follows from Theorem 2.

The condition is also sufficient. Let P be any point of M. I shall show that M is locally separated at P. If ϵ is any positive number, let C_1, C_2, C_3, and C_4 be circles with centers at P with radii $\epsilon/4$, $\epsilon/2$, $3\epsilon/4$, and ϵ, respectively. On a radius of C_4, let the intersections with the circles $C_i(i=1, 2, 3, 4)$ occur in the order *Plkji*, and on the radius diametrically opposite let a and b be

* If "region" be replaced by "sphere," the definition of simple chain and Theorem 10 as given on pp. 134–135 of R. L. Moore's *Foundations of plane analysis situs* (these Transactions, vol. 17 (1916), pp. 131–164) will suffice for reference here.

† It was shown by Urysohn (loc. cit., vol. 8, p. 355) that if M is an F_σ of dimension $<n-1$ in $E_n(n>1)$ and D is a connected domain, then $D-D\times M$ is strongly connected. (As a matter of fact, as I have pointed out in my paper *Concerning a theorem of J. R. Kline*, not yet published, $D-D\times M$ is *arcwise* connected.) Theorem 4a shows that for the case $n=2$ the restriction that M be an F_σ is unnecessary.

the intersections with C_1 and C_2, respectively. On C_1 let m, n, o be points in the order (counter-clockwise) *lmnoa*. Extend the radius Pn of C_1 to meet C_2 at c; and the radius Pm to meet C_2 and C_3 at d and q, respectively. On C_2 let f and e be selected so that the order (counter-clockwise) *kfedcb* is obtained. Let a radius of C_4 through f cut C_3 and C_4 in g and h, respectively, and a radius through e meet C_3 in s. Denoting straight line intervals by brackets, and arcs of circles by parentheses, define simple closed curves J_1 and J_2 as follows:

$$J_1 = [Pab] + (bcdef) + [fgh] + (ih) + [ijkl] + (lmno) + [oP],$$

(ih) being so chosen that J_1 encloses points of $[nc]$;

$$J_2 = [Pmdq] + (qjgs) + [se] + (efkbc) + [cnP].$$

That arc jg on C_3 which does not contain s forms, with the portion $[ji]+(ih)+[hg]$ of J_1, a simple closed curve J_3. Let Q_1 be a point interior to J_3, not belonging to M. As Q_1 is interior to J_1, there exists a continuum K_1 containing Q_1 and P, and lying, except for P, wholly interior to J_1, and containing, except for P, only points of E_2-M. Let

$$J_4 = [fg] + (gs) + [se] + (ef),$$

where (gs) does not contain j, and (ef) does not contain k, and let Q_2 be a point of E_2-M interior to J_4. As Q_2 is also interior to J_2, there exists a continuum K_2 containing Q_2 and P, lying, except for P, wholly interior to J_2, and containing, except for P, only points of E_2-M.

The continuum K_1 contains a continuum T_1 which lies wholly within or on the boundary of the annular domain bounded by C_1 and C_3, and contains points on both C_1 and C_3.* The points of T_1 on C_3 lie on the arc jg of J_3, and the points of T_1 on C_1 are on that arc ao of C_1 which does not contain n. The continuum K_2 contains a continuum T_2 which lies wholly within or on the simple closed curve J_5 defined as follows:

$$J_5 = [fg] + (gjq) + [qdm] + (mn) + [nc] + (cbkf),$$

where (mn) does not contain a, and such that T_2 has points on both $[fg]$ and (mn).

The continuum $T=T_1+T_2$ does not contain P and hence is a subset of E_2-M. If D is that complementary domain of T determined by P, then D is a region whose boundary is a subset of T, and such that D contains no point on or exterior to C_4. The diameter of D is less than ϵ, and hence M is locally separated at P.

* Cf. Anna M. Mullikin, *Certain theorems relating to plane connected point sets*, these Transactions, vol. 24 (1922), pp. 144–162, Theorem 1.

If a set M in E_2 is accessible from all sides, then by Theorem 5 it is zero-dimensional; and hence, by Theorem 2, E_2 is accessible from all sides with respect to M and, by Theorem 3, E_2-M is strongly connected im kleinen. Hence the following corollary:

COROLLARY. *If a set M in E_2 is accessible from all sides, then E_2 is accessible from all sides with respect to M, and E_2-M is strongly connected im kleinen.*

Thus in E_2 accessibility from all sides of a set, and accessibility of E_2 from all sides with respect to that set, are equivalent properties.

That ordinary accessibility of points of M, in the sense that if P is a point of M and Q a point of E_2-M there exists a continuum K containing P and Q and such that $K\times M=P$, is not sufficient, even where M is punctiform and totally disconnected, to insure the zero-dimensionality of M is shown by the example of a quasi-connected point set in my paper *A set which has no true quasi-components and which becomes connected upon the addition of a single point.** The set described in this paper is accessible by arcs.

That in the definition of accessibility from all sides used above, arcs cannot be employed instead of unrestricted continua will be shown by the example given in §2 of a zero-dimensional set which is not simply locally separated.

2

It is known† that the set, I_2, of all points in E_2 both of whose coördinates are irrational is homeomorphic with the set, I_1, of points in E_1 whose abscissas are irrational. Sierpinski has shown that zero-dimensional sets are homeomorphic with linear sets and hence with subsets of I_1.‡ Hence every zero-dimensional set is homeomorphic with a subset of I_2. This suggests the question: Is a zero-dimensional set in E_2 isotopic with a subset of I_2? It is to be noticed, first, that any subset of I_2 is simply locally separated as defined above, and that any set isotopic with a subset of I_2 must accordingly be simply locally separated. Hence I shall first show that the above question cannot be answered affirmatively by giving an example of a zero-dimensional set in E_2 which is not simply locally separated at any point.

---

* Bulletin of the American Mathematical Society, vol. 33 (1927), pp. 423–427.

† Cf. M. Fréchet, *Les dimensions d'un ensemble abstrait*, Mathematische Annalen, vol. 68 (1910), pp. 145–168. See especially p. 154.

‡ Cf. Menger, *Bericht über die Dimensionstheorie*, loc. cit., pp. 125–126, and references given therein.

Consider, first, a set M constructed as follows: Let T be a unit square, and for each positive integer n divide T into n^2 equal squares, letting T_n denote the corresponding set of n^2 squares. For each square, t, of T_n, construct a continuum $K(t)$ of which every subcontinuum is indecomposable* and which lies wholly within t except that it contains four points on t, these being the mid-points of the four sides of t. For each n let M_n denote the set of points obtained by adding together the point sets $K(t)$ for all squares t of T_n. Let $M=\sum_1^\infty M_i$. This set, M, was first constructed by R. L. Moore† as an example of a connected and connected im kleinen point set which contains no arc. The point set which I wish now to consider is the complement of M within and on T. Denote this set by N.

The point set N is a zero-dimensional G_δ‡ and is everywhere dense in T and its interior. For if P is a point of N within T and C is a circle with center at P, there exists a positive integer n such that not only does a square t of T_n, within or on which P lies, lie wholly within C, but also all those squares of T_n adjacent to t. The eight continua K_t constructed relative to T_n in these adjacent squares form a continuum F. That complementary domain D of F which contains P is a region which lies wholly interior to C and whose boundary contains no point of N, being a subset of F and hence of M. Then N is locally separated at P. In case P is on T it can be shown in a similar way that N is locally separated at P. That N is everywhere dense in T and its interior is evident since every arc interior to T contains points of N. That N is a G_δ is evident since M is an F_σ. But it is obvious that N is not simply locally separated at any point and consequently N is not isotopic with I_2.

By an extension of this example there can be obtained an example of a zero-dimensional set which is dense everywhere in E_2, but which is not isotopic with I_2. This comment is of interest, perhaps, in that it relates to the analogue, for punctiform uncountable sets, of a theorem first given by Fréchet§ and later by Urysohn‖ to the effect that all denumerable sets dense in E_n are isotopic. If there exists any analogue of this theorem for uncountable sets, the example just indicated shows that zero-dimensionality is not a sufficient condition for isotopism of such sets.

---

* Cf. B. Knaster, *Un continu dont tout sous-continu est indécomposable*, Fundamenta Mathematicae, vol. 3 (1922), pp. 247–286.

† R. L. Moore, *A connected and regular point set which contains no arc*, Bulletin of the American Mathematical Society, vol. 32 (1926), pp. 331–332.

‡ A G_δ is a set of points common to a denumerable infinity of open sets.

§ M. Fréchet, loc. cit., p. 159.

‖ P. Urysohn, *Sur les multiplicités Cantoriennes*, loc. cit., vol. 7, pp. 83 ff. Urysohn states that Fréchet's proof seems to him insufficient.

I shall now proceed to establish a condition which characterizes those sets which are isotopic with subsets of I_2.

THEOREM 6. *Let M be a point set in E_2. Then in order that M should be simply locally separated, it is necessary and sufficient that E_2 be arcwise accessible from all sides with respect to M.*

The condition is necessary. Let M be a simple locally separated set in E_2, and let P be any point of E_2. Let J be any simple closed curve containing P. Let D be the bounded domain complementary to J, and let Q be any point of D.

There exists a simple closed curve K which contains P and Q, and which lies, except for P, wholly in D. Denote the interior of K by R. There exists an arc t which has P and Q as end points, and which lies, except for these two points, entirely in R.

Denote the point set $M \times [K-(P+Q)]$ by M_1.

Since M is simply locally separated, there exists, by Theorem 1a, a sequence, G, of Jordan regions, which covers M in the Vitali sense and has property H, and such that the boundaries of the regions of G contain no points of M.

If x is a point of M_1, there exists a region g_x of the collection G which contains x and such that g_x' lies wholly in D and contains no point of t. The set of all points which lie in regions of the type g_x together with their boundaries denote by T. The set of points $T+K$ is a continuum C. Furthermore, C is a continuous curve. To show this, I shall employ Sierpinski's characterization* of a continuous curve; i.e., a bounded continuum N is a continuous curve provided that for every positive number e, N is the sum of a finite collection of continua each of which is of diameter less than e.

Let e be any positive number. Then K, being a continuous curve, is the sum of a finite collection of continua, $C_1, C_2, \cdots, C_n$, each of which is of diameter less than $e/4$. Since G covers M in the Vitali sense, there is only a finite number of regions of G of diameter $>e/4$. Hence those regions of G which constitute part of T and are of diameter $>e/4$ are finite in number, and as a Jordan region together with its boundary forms a continuous curve, each of them together with its boundary is the sum of a finite number of continua of diameters less than e; denote the set of all such continua by G_1. If g_x is a region which constitutes part of T and is of diameter less than $e/4$, and if C_i $(1 \leqq i \leqq n)$ is a part of K which contains a point of g_x', then C_i,

* W. Sierpinski, *Sur une condition pour qu'un continu soit une courbe jordanienne*, Fundamenta Mathematicae, vol. 1 (1920), pp. 44–60.

together with all such regions g_z and their boundaries forms a continuum C_1'. That C_1' is of diameter less than e is obvious. Therefore C is the sum of a finite collection of continua, viz., C_1', C_2', $\cdots$, C_n', and the continua of the collection G_1, all of which are of diameter less than e, and hence is a continuous curve.

If z is an interior point of t, denote by D_z that complementary domain of C determined by z. Denote the outer boundary of D_z by B. By a theorem due to R. L. Moore,* B is a simple closed curve. That B contains P and Q is easily seen, and clearly B lies wholly in D, except for P, and contains no point of M except possibly P and Q. As B is the sum of two arcs whose end points are P and Q, the condition stated in the theorem is proved necessary.

To show that the condition stated in the theorem is sufficient a modification of the proof of Theorem 5 may be employed.

THEOREM 6a. *In order that a point set in E_2 should be simply locally separated it is necessary and sufficient that it be arcwise accessible from all sides.*

COROLLARY. *If a set M in E_2 is arcwise accessible from all sides, then E_2 is arcwise accessible from all sides with respect to M, and E_2-M is arcwise connected im kleinen; and if P and Q are points of a connected domain D in E_2 there is an arc from P to Q which lies wholly in D and contains no point of M except possibly P and Q.*

The proof of the latter part of this corollary can be obtained by use of the simple chain idea (see proof of Theorem 4a), the arcwise accessibility of E_2 with respect to M, and of the fact that if C is a circle enclosing two points A and B there exists a simple closed curve within C which contains A and encloses B.

THEOREM 7. *In E_2, let M be a point set and let I_2 denote the set of all points both of whose coördinates are irrational. Then in order that M should be isotopic with a subset of I_2, it is necessary and sufficient that it be simply locally separated.*

That the condition stated in the theorem is necessary is obvious, since I_2 is itself simply locally separated.

To prove the condition sufficient, I shall proceed by methods based on R. L. Moore's *Concerning a set of postulates for plane analysis situs.*† I shall therefore first state a series of lemmas which are adaptations, as signified

* *Concerning continuous curves in the plane*, loc. cit., Theorem 4. That the boundary of D_z is itself a continuous curve is demonstrated by Moore in the proof of Theorem 4.

† These Transactions, vol. 20 (1919), pp. 169–178.

in each case, of Moore's theorems B-G. It will be understood that when a theorem is designated by letter it is one of Moore's theorems, and that when a theorem is designated by number it refers to a theorem in the present paper. Wherever a proof is not indicated it will be understood that the proof is only a slight modification of Moore's proof with the use of Theorem 6.

LEMMA 3 (Adaptation of Theorem B). *If J and L are two simple closed curves and A and B are two distinct points of $J\times(E_2-M)$ each of which is either not on L at all or on some segment that is common to J and L, then there exists an arc from A to B which lies entirely in E_2-M and, except for its end points, within J, and has not more than a finite number of points in common with L.*

The proof of Lemma 3 is similar to that of Theorem B, except that use is made of Theorem 6 (a) in establishing the existence of the arc AB in E_2-M, in the first sentence of the proof as given by Moore, and (b) in establishing the existence of the arcs $A_nZ_nB_n$ of lines 10–12, page 172, so that these have, in addition to the properties outlined there, the further property that they lie in E_2-M.

LEMMA 4 (Adaptation of Theorem C). *If the closed curve g lies in E_2-M and has only a finite number of points in common with the closed curve $ABCDA$ and does not contain A, B, C, or D, then the interior of $ABCDA$ can be divided by double ruling* such that* (1) *the arcs of this ruling lie in E_2-M,* (2) *the arcs of one of its single rulings are parallel to AB and CD and those of the other are parallel to AD and BC, and* (3) *the subdivisions of $ABCDA$ made by this ruling are such that the interior of each one of them is either wholly within or wholly without g.*

LEMMA 5 (Adaptation of Theorem D). *If $ABCDA$ is a simple closed curve and G is a set of simple closed curves which lie wholly in E_2-M and each point on or within $ABCDA$ is within some curve of the set G, then the interior of $ABCDA$ can be divided by a double ruling such that* (1) *the arcs of this ruling lie wholly in E_2-M,* (2) *the arcs of one of its single rulings are parallel to AB and CD and those of the other are parallel to AD and BC, and* (3) *the subdivisions of the interior of $ABCDA$ formed by these rulings are such that each lies within some curve of the set G.*

LEMMA 6 (Adaptation of Theorem E). *If $ABCDA$ is a simple closed curve there exist two sets of arcs, α_1 and α_2, such that* (1) *each arc of α_1 lies wholly within $ABCDA$ except that its end points are on AB and CD,* (2) *each arc of*

* For definitions see Moore's paper, loc. cit.

α_2 lies wholly within ABCDA except that its end points are on BC and DA, (3) *$\alpha_i (i=1, 2)$ is the sum of two collections of arcs, a_i and b_i, such that* (i) *if d is an arc of b_i, then d is the sequential limiting set of a sequence of arcs of the set a_i and a_i is a denumerable set,* (ii) *the arcs of a_i contain no points of M,* (4) *each point on ABCDA, with the exception of A, B, C, and D, is an end point of either just one arc of α_1 or just one arc of α_2,* (5) *through each point within ABCDA there is just one arc of α_1 and just one arc of α_2,* (6) *each arc of α_1 has just one point in common with each arc of α_2.*

The proof of Lemma 6 is the same as that of Theorem E, the curves in each set β_n being selected in $E_2 - M$, however.

LEMMA 7 (Adaptation of Theorem F). *There exists a denumerably infinite sequence of simple closed curves $J_1, J_2, J_3, \cdots$ such that every point of E_2 lies within at least one of them and such that for every n, J_{n+1} encloses J_n, and J_n lies wholly in $E_2 - M$.*

The proof of Lemma 7 is either a modification of the proof of Theorem F or an application of Theorem 6 to a series of annular domains.

LEMMA 8 (Adaptation of Theorem G). *There exist in E_2 two sets, G_1 and G_2, of open curves such that* (1) *through each point there is just one curve of G_1 and just one curve of G_2,* (2) *each curve of G_1 has just one point in common with each curve of G_2,* (3) *$G_i (i=1, 2)$ consists of two sets of open curves G_{i1} and G_{i2} such that* (a) *no curve of the set G_{i1} contains a point of M,* (b) *the set G_{i1} is denumerable,* (c) *every curve of the set G_{i2} is the sequential limiting set of a sequence of curves of the set G_{i1}.*

To establish the sufficiency of the condition stated in Theorem 7, proceed as follows on the basis of Lemma 8: Let g_1 and g_2 be particular open curves of the collection G_{11} and G_{21}, respectively. Between the points of g_1 and the points of the x-axis there is a one-to-one continuous correspondence in which the intersections of g_1 with curves of the set G_{21} correspond to points on the x-axis whose abscissas are rational and the intersections of g_1 with curves of the set G_{22} correspond to points on the x-axis whose abscissas are irrational, and in which the intersection of g_1 with g_2 corresponds to the point of the x-axis whose abscissa is zero. Similarly, between the points of g_2 and the points of the y-axis there exists a one-to-one continuous correspondence in which the intersections of g_2 with the curves of the collection G_{11} correspond to the points of the y-axis whose ordinates are rational and the intersections with the curves of the collection G_{12} correspond to the points whose ordinates are irrational and in which the intersection of g_1 and g_2 corresponds to the point whose ordinate is zero.

Let Σ_1 denote the ordinary cartesian system of coördinates, and let Σ_2 denote a system of coördinates defined as follows: If P is any point of E_2, the intersection of that curve of the collection G_2 which contains P with g_1 corresponds, in the correspondence outlined above, to a point of the x-axis whose abscissa is, say, x_2; and the intersection of that curve of G_1 which contains P with g_2 corresponds to a point of the y-axis whose ordinate is y_2; then the coördinates of P in the system Σ_2 will be (x_2, y_2). Then there exists a one-to-one continuous transformation of E_2 into itself in which two points correspond if and only if their respective coördinates in the two systems Σ_1 and Σ_2 are identical, each to each. Since points of M lie only on curves of the sets G_{21} and G_{22}, it is evident that these points correspond to points both of whose coördinates are irrational in the system Σ_1; i.e., the points of M correspond to a subset of I_2.

As a consequence of Theorems 6a and 7 we have

THEOREM 7a. *In order that a set of points in E_2 should be isotopic with a subset of I_2, it is necessary and sufficient that it be arcwise accessible from all sides.*

THEOREM 8. *In E_2, let M be a punctiform F_σ;* then E_2 is arcwise accessible from all sides with respect to M.*

In my paper *Concerning a theorem of J. R. Kline*, I have shown that if, in E_2, D is any connected domain and N is a punctiform F_σ, then the set $D-D\times N$ is arcwise connected. Accordingly, if J is any simple closed curve and P is a point of J, and Q is a point of the region, R, bounded by J, it is easy to prove that there exists an arc from P to Q which lies, except for P, and possibly Q, wholly in the set $D\times(E_2-M)$.

As a consequence of Theorems 6 and 8 we have

THEOREM 9. *In E_2 every punctiform F_σ is simply locally separated.*

As a consequence of Theorems 7 and 9 we have

THEOREM 10. *In E_2 every punctiform F_σ is isotopic with some subset of I_2, the set of all points both of whose coördinates are irrational.*

3. SOME PROBLEMS

1. Under what conditions is a zero-dimensional set in E_n $(n>2)$ isotopic with a subset of I_n, the set of points all of whose coördinates are irrational?

---

* A set is an F_σ if it is the sum of a denumerable collection of closed sets. Cf. F. Hausdorff, *Grundzüge der Mengenlehre*, Leipzig, 1914, pp. 304 ff.

2. If M is a punctiform F_σ in E_n $(n>2)$, is M isotopic with some subset of I_n?

3. When is a zero-dimensional set M in E_n $(n>2)$ arcwise accessible from all sides, and when is E_n arcwise accessible from all sides with respect to M? Are the two properties equivalent?

4. If M is a zero-dimensional set in E_n $(n>2)$ under what conditions is the complement of M arcwise connected? More generally, if D is a connected domain, when is $D-D\times M$ arcwise connected; when are any two points of D, say P and Q, the end points of an arc of D which contains no point of M, except possibly P and Q?

Problems 3 and 4 are obviously closely related, for it is easy to show that if any set M (zero-dimensional or not) is such that E_n is arcwise accessible from all sides with respect to M, and D is a connected domain and P and Q are points of D, then D contains an arc from P to Q which contains no point of M except possibly P and Q.

5. What is the maximum dimension of a point set which is accessible from all sides in E_n?

The solution of Problem 5 for $n=2$ is obviously given by Theorem 5.

6. Let M be an arbitrary point set and D a connected domain in E_n. Is there a more general condition than that given by Urysohn (see footnote accompanying Theorem 4a) sufficient to ensure the *strong* connectivity of $D-D\times M$?

UNIVERSITY OF MICHIGAN,
ANN ARBOR, MICH.

ON THE LINEAR PARTIAL q-DIFFERENCE EQUATION OF GENERAL TYPE*

BY

C. RAYMOND ADAMS

Introduction. In a paper published some time ago† the author defined the linear partial q-difference equation and examined a class of such equations. The general methods employed were those introduced by Birkhoff‡ in studying the linear ordinary difference equation. It is our purpose here to investigate the general problem of the partial q-difference equation, to which the methods of Birkhoff, based primarily on iteration, are not applicable. We propose to establish existence theorems by direct convergence proofs.

The equation considered is§

$$\sum_{i,j=0}^{n} a_{ij}(x,y)f(q^i x, r^j y) = 0, \tag{1}$$

in which the $a_{ij}(x, y)$ are known functions of the complex variables x and y, analytic at the place $(0, 0)$; q and r are constants, not zero and of moduli $\neq 1$; and $f(x, y)$ is the function to be determined. We may write

$$a_{ij}(x,y) = a_{ij00} + a_{ij10}x + a_{ij01}y + \cdots \text{ for } |x| < R_1, \quad |y| < R_2 \quad (i,j = 0,1,\cdots,n), \tag{2}$$

and without loss of generality we may and do assume $|q|$ and $|r|$ both greater than 1.

For convenience in notation we define

$$\begin{aligned} g(\rho,\sigma,k,l,m,p) &= \sum_{i,j=0}^{n} a_{ijmp}\rho^i q^{ki}\sigma^j r^{lj}, \\ g_{s0}(\rho,\sigma,k,l,m,p) &= \sum_{i,j=0}^{n} i^s a_{ijmp}\rho^i q^{ki}\sigma^j r^{lj}, \\ g_{0s}(\rho,\sigma,k,l,m,p) &= \sum_{i,j=0}^{n} j^s a_{ijmp}\rho^i q^{ki}\sigma^j r^{lj} \quad (s = 1,2,\cdots). \end{aligned} \tag{3}$$

---

* Presented to the Society, September 7, 1928; received by the editors August 24, 1928.

† Adams, *The general theory of a class of linear partial q-difference equations*, these Transactions, vol. 26 (1924), pp. 283–312; this paper will hereafter be referred to as I.

‡ Birkhoff, *General theory of linear difference equations*, these Transactions, vol. 12 (1911), pp. 243–284.

§ It is clearly possible to write equation (B) of I in this form.

Then the algebraic equation

$$g(\rho,\sigma,0,0,0,0) = 0 \tag{4}$$

is called the *characteristic equation of* (1) *for* (0, 0). We speak of a pair of values ρ_0, σ_0 satisfying (4) as the *characteristic number-pair* (ρ_0, σ_0); it is said to be *simple* if $g(\rho, \sigma_0, 0, 0, 0, 0) = 0$ has ρ_0 as a simple root and if $g(\rho_0, \sigma, 0, 0, 0, 0) = 0$ has σ_0 as a simple root.

We term (ρ_0, σ_0) a *regular* characteristic number-pair when the following assumptions are fulfilled: (a) both ρ_0 and σ_0 are finite and different from zero; (b) (ρ_0, σ_0) is simple and satisfies none of the equations

$$g(\rho,\sigma,k,l,0,0) = 0, \tag{5}$$

where k and l are positive integers or zero but not both zero; (c) a_{nn00} is not zero; (d) ρ_0 satisfies none of the equations

$$\sum_{i=0}^{n} a_{in00}\rho^i q^{ki} = 0 \qquad (k = 0,1, \cdots, K-1);$$

(e) σ_0 satisfies none of the equations

$$\sum_{j=0}^{n} a_{nj00}\sigma^j r^{lj} = 0 \qquad (l = 0,1, \cdots, L-1).$$

K and L are integers presently to be determined.

We direct our attention first, in §§1, 2, to the case of a regular characteristic number-pair, exhibiting a corresponding series that formally satisfies (1) and by proof of convergence demonstrating the existence of a solution analytic in the vicinity of (0, 0). In §§3, 4 we examine, with similar results, the case of an irregular characteristic number-pair whose irregularity is due to the failure of assumption (b) alone. §5 is devoted to the cases of other irregular characteristic number-pairs. We conclude in §6 with a brief consideration of the non-homogeneous equation obtained by adding to the left-hand member of (1) a known function.

It is of interest and importance to observe (i) that §§1, 2 provide a simple and brief derivation of the fundamental existence theorem of §2 of I; (ii) that if the coefficient functions $a_{ij}(x, y)$ are analytic at the place (∞, ∞), or at the place $(0, \infty)$, or at the place $(\infty, 0)$, the situation with regard to solutions in the vicinity of this place is entirely parallel to that described in detail here; and (iii) that the methods employed here are applicable at once to the linear partial q-difference equation of general type in N independent variables.

1. Regular characteristic number-pair; formal series. Corresponding to a regular characteristic number-pair (ρ_0, σ_0) there exists a formal solution of (1),

$$s(x,y) = x^a y^b (s_{00} + s_{10}x + s_{01}y + \cdots), \tag{6}$$

where*

$$a = \frac{\log \rho_0}{\log q} \text{ and } b = \frac{\log \sigma_0}{\log r}.$$

The coefficients in (6) may be calculated by substitution of the series in (1); s_{00} is arbitrary, but the subsequent coefficients are determined uniquely in terms of it by the relation

$$s_{kl} = -\frac{\sum_{i=0}^{k-1} s_{il} g(\rho_0,\sigma_0,i,l,k-i,0) + \sum_{i=0}^{k}\sum_{j=0}^{l-1} s_{ij} g(\rho_0,\sigma_0,i,j,k-i,l-j)}{g(\rho_0,\sigma_0,k,l,0,0)}. \tag{7}$$

It may be remarked that assumptions (c), (d), and (e) are not needed for this section.

2. Regular characteristic number-pair; existence of an analytic solution. For convenience we assume that the associated radii of convergence of the series (2) are all greater than 1. It can readily be shown, as in the case of the linear ordinary q-difference equation, that this hypothesis is no real restriction.†

We seek to prove the convergence of the series (6) by finding a convergent dominant series. Let us first examine the denominator of (7) with a view to establishing the inequality

$$|g(\rho_0,\sigma_0,k,l,0,0)| > C\,|q^{n(a+k)} r^{n(b+l)}| \qquad (C>0) \tag{8}$$

for $k, l = 0, 1, \cdots$ but not both zero. After the factor $|q^{n(a+k)} r^{n(b+l)}|$ is removed from $|g(\rho_0, \sigma_0, k, l, 0, 0)|$, the remaining factor is

$$\left| a_{nn00} + \frac{a_{n,n-1,00}}{r^{b+l}} + \cdots + \frac{a_{n000}}{r^{n(b+l)}} + \frac{a_{n-1,n00}}{q^{a+k}} + \frac{a_{n-1,n-1,00}}{q^{a+k} r^{b+l}} + \cdots \right.$$
$$\left. + \frac{a_{n-1,000}}{q^{a+b} r^{n(b+l)}} + \cdots + \frac{a_{0n00}}{q^{n(a+k)}} + \frac{a_{0,n-1,00}}{q^{n(a+k)} r^{b+l}} + \cdots + \frac{a_{0000}}{q^{n(a+k)} r^{n(b+l)}} \right|. \tag{9}$$

* Throughout the paper we take, for definiteness, those determinations of $\log x$, $\log y$, $\log \rho_0$, $\log \sigma_0$, $\log q$, and $\log r$ in which the coefficient of $(-1)^{1/2}$ is positive or zero and less than 2π.

† Adams, *On the linear ordinary q-difference equation*, Annals of Mathematics, (2), vol. 30 (1929), No. 2; we shall refer to this paper as II.

Since a_{nn00} is not zero, a value of l, which we denote by L, can be found to make

$$\left|\frac{a_{n,n-1,00}}{r^{b+l}}+\cdots+\frac{a_{n000}}{r^{n(b+l)}}\right|<\frac{|a_{nn00}|}{n+2} \quad \text{for} \quad l\geqq L.$$

Then we have

$$\left|a_{in00}+\frac{a_{i,n-1,00}}{r^{b+l}}+\cdots+\frac{a_{i000}}{r^{n(b+l)}}\right|\leqq|a_{in00}|+\left|\frac{a_{i,n-1,00}}{r^{b+L}}\right|$$
$$+\cdots+\left|\frac{a_{i000}}{r^{n(b+L)}}\right| \text{ for } l\geqq L \qquad (i=0,1,\cdots,n-1),$$

and a value of k, which we denote by K, can be found to yield the inequality

$$\frac{|a_{in00}|+\left|\frac{a_{i,n-1,00}}{r^{b+L}}\right|+\cdots+\left|\frac{a_{i000}}{r^{n(b+L)}}\right|}{|q^{(n-i)(a+k)}|}<\frac{|a_{nn00}|}{n+2} \quad \text{for} \quad k\geqq K,\ l\geqq L$$
$$(i=0,1,\cdots,n-1).$$

Thus for $k\geqq K, l\geqq L$ the quantity (9) is greater than $|a_{nn00}|/(n+2)$.

For any particular $l<L$ we can, by virtue of assumption (e), choose k_l so large that the sum of terms containing q in (9) is in absolute value less than

$$\frac{1}{2}\left|a_{nn00}+\frac{a_{n,n-1,00}}{r^{b+l}}+\cdots+\frac{a_{n000}}{r^{n(b+l)}}\right| \quad \text{for} \quad k\geqq k_l, \tag{10}$$

and the quantity (9) is then greater than (10).

For any particular $k<K$ we can, because of assumption (d), choose l_k so large that the sum of terms containing r in (9) is in absolute value less than

$$\frac{1}{2}\left|a_{nn00}+\frac{a_{n-1,n00}}{q^{a+k}}+\cdots+\frac{a_{0n00}}{q^{n(a+k)}}\right| \quad \text{for} \quad l\geqq l_k, \tag{11}$$

and the quantity (9) is then greater than (11).

Now there are only a finite number of pairs of values k, l not included in the three classes $k\geqq K,\ l\geqq L$; $k<K,\ l\geqq l_k$; $l<L,\ k\geqq k_l$. For each of these pairs, by assumption (b), the quantity (9) has a positive value C_{kl}. If then we take for C the smallest of the constants

$$\frac{|a_{nn00}|}{n+2};\ (10) \text{ for } l=0,1,\cdots,L-1;\ (11) \text{ for } k=0,1,\cdots,K-1;\ C_{kl},$$

the inequality (8) will be satisfied.

Turning to the numerator of (7) and recalling that the series (2) converge for $x=y=1$, we have

$$|a_{ijmp}| < M' \qquad (i,j = 0,1,\cdots,n; m,p = 0,1,\cdots),$$

M' being a suitable positive number. Let the largest of the quantities $|q^{ia}r^{jb}|(i,j=0,1,\cdots,n)$ be denoted by M''. Since $|q|$ and $|r|$ are greater than 1, we then have

$$|s_{kl}| < \frac{1}{C|q^{n(a+k)}r^{n(b+l)}|}\left[(n+1)^2M'M''|q^{nk}r^{nl}|\left(\sum_{i=0}^{k-1}|s_{il}| + \sum_{i=0}^{k}(|s_{i0}| + |s_{i1}| + \cdots + |s_{i,l-1}|)\right)\right].$$

Defining $\bar{s}_{00}$ as $|s_{00}|$ and setting

$$M = \frac{(n+1)^2M'M''}{C|q^{na}r^{nb}|},$$

we obtain the following bound for s_{kl}:

$$\bar{s}_{kl} = M\left[\sum_{i=0}^{k-1}(\bar{s}_{i0} + \bar{s}_{i1} + \cdots + \bar{s}_{il}) + \bar{s}_{k0} + \bar{s}_{k1} + \cdots + \bar{s}_{k,l-1}\right], \tag{12}$$

k and l being any positive integers or zero but not both zero. From this formula follows at once the recurrence relation

$$\bar{s}_{kl} = M\sum_{i=0}^{k-1}\bar{s}_{il} + (M+1)\bar{s}_{k,l-1}, \tag{13}$$

in which $\bar{s}_{k,-1}$ and $\bar{s}_{-1,l}$ are to be interpreted as zero.

It is possible to show directly that the series

$$\sum_{k,l=0}^{\infty}\bar{s}_{kl}x^ky^l \tag{14}$$

has associated radii of convergence greater than zero. For simplicity and for convenience in our later work, however, we prefer to establish this fact indirectly, by proving the same property for the dominant series

$$\sum_{k,l=0}^{\infty}S_{kl}x^ky^l, \tag{15}$$

where $S_{00}=\bar{s}_{00}$ and S_{kl} is determined by the recurrence formula

$$S_{kl} = (M+1)\left[\sum_{i=0}^{k-1}S_{il} + S_{k,l-1}\right] \qquad (S_{k,-1} = S_{-1,l} = 0). \tag{16}$$

The basis of our proof is the rather obvious

LEMMA. *A sufficient condition that a power series in two variables with positive coefficients have associated radii of convergence greater than zero is that the simple power series which it becomes upon setting $y=x$ have a positive radius of convergence.*

From (16) we have*

$$S_{kl} \leqq (M+1)(S_{k-1,l}+S_{k,l-1})+S_{k-1,l} \leqq (M+2)(S_{k-1,l}+S_{k,l-1}).$$

The series (15) is therefore dominated by

$$T(x,y)=\sum_{k,l=0}^{\infty} T_{kl}x^k y^l, \tag{17}$$

in which $T_{00}=S_{00}$ and T_{kl} is given by the relation

$$T_{kl}=(M+2)(T_{k-1,l}+T_{k,l-1}) \quad (T_{k,-1}=T_{-1,l}=0).$$

In the simple power series

$$T(x,x)=\sum_{\nu=0}^{\infty} T_\nu x^\nu,$$

the coefficients are

$$T_\nu=\sum_{k+l=\nu} T_{kl}=2(M+2)\sum_{k+l=\nu-1} T_{kl}=2(M+2)T_{\nu-1}.$$

Hence $T(x, x)$ converges for $|x|<1/[2(M+2)]$ and the series (17), (15), (14), and (6) all have associated radii of convergence that are positive. Thus we obtain

THEOREM I. *Corresponding to each regular characteristic number-pair (ρ_0, σ_0) there exists a solution of the equation* (1) *which is analytic in the vicinity of the place* (0, 0) *and is expressed there by the series* (6).

It may be added that *if $a_{in}(x, y)$ $[a_{ni}(x, y)]$ vanishes identically for $i=0, 1, \cdots, n-1$, the region of analyticity of the solutions can be extended away from the origin in the x-plane [y-plane] by repeated use of the equation* (1) *itself; if both of these conditions are satisfied, each of the series* (6) *converges out to the nearest place which is a singularity of the coefficient functions $a_{ij}(x, y)$ or a zero of $a_{nn}(x, y)$ and represents an analytic solution of* (1) *thus far out from* (0, 0); *if, further, the a_{ij} (x, y) are all rational functions, these solutions analytic*

* The proof of convergence of (15), as originally constructed by the author, is based on a comparison of $\sum_{k+l=\nu} S_{kl}$ with $S_{00}(1+1+1)^\nu$, by which it is dominated. The particularly simple discussion given in this paragraph follows the lines of a suggestion by J. D. Tamarkin.

near (0, 0) *are analytic except for poles over the entire finite x- and y-planes, except perhaps for* $x=0$ *and* $y=0$.

3. **Irregular characteristic number-pair, the irregularity arising from failure of assumption (b) only; formal series.** Let us consider first a characteristic number-pair (ρ_0, σ_0) which is of multiplicity $\mu(\geqq 1)$ in ρ(i.e., $g(\rho, \sigma_0, 0, 0, 0, 0)=0$ has ρ_0 as a root of multiplicity μ) and which satisfies exactly $\nu(\geqq 0)$ of the equations (5), each equation being counted a number of times equal to the multiplicity in ρ of the solution (ρ_0, σ_0). To this number-pair there correspond, whether (ρ_0, σ_0) is simple or multiple in σ, μ formal solutions of (1),

$$(18)\quad s_i(x,y)=x^a y^b[(\text{P.S.})+t(\text{P.S.})+\cdots+t^{\nu+i-1}(\text{P.S.})]\quad (i=1,2,\cdots,\mu),$$

where

$$a=\frac{\log \rho_0}{\log q},\qquad b=\frac{\log \sigma_0}{\log r},\qquad t=\frac{\log x}{\log q},$$

and (P.S.) is used in a generic sense to denote a power series in x and y. The coefficients in (18) may be calculated by substitution of the series in(1). In the power series in $s_i(x, y)$ there are in general* $\nu+i$ coefficients whose values may be assigned arbitrarily; all of the other coefficients are then uniquely determined.

It is now quite evident from symmetry that if (ρ_0, σ_0) is of multiplicity μ in σ and satisfies exactly ν of the equations (5), each equation being counted a number of times equal to the multiplicity in σ of the solution (ρ_0, σ_0), there correspond, whether (ρ_0, σ_0) is simple or multiple in ρ, μ formal solutions precisely like (18) except that t is replaced by

$$\tau=\frac{\log y}{\log r}.$$

The two foregoing paragraphs suggest an interesting question. One might rather naturally expect, if ν is zero and if (ρ_0, σ_0) is of multiplicity μ_1 in ρ and of multiplicity μ_2 in σ, that there should be $\mu_1 \cdot \mu_2$ solutions corresponding to this number-pair. Only $\mu_1+\mu_2-1$ independent solutions, however, have been exhibited here. Are there other solutions, and if so of what form are they?

A second point of special interest arises when ν is $\geqq 1$. Let us suppose, for simplicity, that (ρ_0, σ_0) is simple; then one might expect only one formal solution corresponding to it. Yet the condition (5) does not favor x and q

* Under special conditions there may be more.

over y and r, and it is fairly clear that there must be a second independent solution like (18) except for the replacement of t by τ. In fact not only is this true but it can readily be verified also that in general there are still other independent formal solutions containing powers of both t and τ,

$$(19)\quad x^a y^b[(\text{P.S.}) + t(\text{P.S.}) + \cdots + t^n(\text{P.S.}) + \tau(\text{P.S.}) + \cdots + \tau^n(\text{P.S.})].$$

If (ρ_0, σ_0) is multiple, the situation is analogous.

As in §1, the assumptions (c), (d), and (e) are not needed for this section.

4. Irregular characteristic number-pair, the irregularity arising from failure of assumption (b) only; existence of analytic solutions. We consider in detail the case in which the equation (1) is satisfied formally by the series (18). Let us examine

$$(20)\quad s_1(x,y) = x^a y^b[(s_{00}^{(\nu)} + s_{10}^{(\nu)}x + s_{01}^{(\nu)}y + \cdots) + t(s_{00}^{(\nu-1)} + s_{10}^{(\nu-1)}x + s_{01}^{(\nu-1)}y + \cdots) + \cdots + t^\nu(s_{00}^{(0)} + s_{10}^{(0)}x + s_{01}^{(0)}y + \cdots)],$$

assuming for the sake of generality that ν is positive; the methods employed are then clearly applicable to $s_i(x, y)$ $(i=2, 3, \cdots, \mu)$. As in §2 we make no real restriction in assuming the associated radii of convergence of all the series (2) greater than 1.

In order to prove the power series in $s_1(x, y)$ convergent in the vicinity of (0, 0) we observe first that the coefficients $s_{kl}^{(m)}$ in these series satisfy the following relations (cf. (3)):

$$(21)\quad \begin{aligned} s_{kl}^{(m)} = -\Bigg\{ &\sum_{i=0}^{k-1} s_{il}^{(m)} g(\rho_0,\sigma_0,i,l,k-i,0) + \sum_{i=0}^{k}\sum_{j=0}^{l-1} s_{ij}^{(m)} g(\rho_0,\sigma_0,i,j,k-i,l-j) \\ &+ \sum_{i=0}^{k}\sum_{j=0}^{l}\Bigg[\binom{\nu-m+1}{1} s_{ij}^{(m-1)} g_{10}(\rho_0,\sigma_0,i,j,k-i,l-j) \\ &+ \binom{\nu-m+2}{2} s_{ij}^{(m-2)} g_{20}(\rho_0,\sigma_0,i,j,k-i,l-j) + \cdots \\ &+ \binom{\nu}{m} s_{ij}^{(0)} g_{m0}(\rho_0,\sigma_0,i,j,k-i,l-j)\Bigg]\Bigg\} \div g(\rho_0,\sigma_0,k,l,0,0) \\ &(m = 0,1,\cdots,\nu). \end{aligned}$$

Whenever $g(\rho_0, \sigma_0, k, l, 0, 0)$ vanishes, the numerator of this quotient is of course also zero. Bounds for the quantities $s_{kl}^{(m)}$ are determined here as they were for the s_{kl} in §2. We let M' and M'' have the same significance as before, let B stand for the largest of the binomial coefficients appearing in any one of the expressions (21), and redefine M as follows:

$$M = \frac{(n+1)^2 n^r B M' M''}{C \left| q^{n a_r n b} \right|}.$$

Thus, defining $\bar{s}_{00}^{(0)}$ as $\left| s_{00}^{(0)} \right|$ and in fact $\bar{s}_{kl}^{(m)}$ whenever $s_{kl}^{(m)}$ is arbitrary as $\left| s_{kl}^{(m)} \right|$, we obtain a bound for (21):

$$\bar{s}_{kl}^{(m)} = M\Big[\sum_{i=0}^{k-1} \bar{s}_{il}^{(m)} + \sum_{i=0}^{k} \sum_{j=0}^{l-1} \bar{s}_{ij}^{(m)} + \sum_{i=0}^{k} \sum_{j=0}^{l} (\bar{s}_{ij}^{(m-1)} + \bar{s}_{ij}^{(m-2)} + \cdots + \bar{s}_{ij}^{(0)}) \Big]. \tag{22}$$

For $m=0$ the right-hand member reduces precisely to that of (12); we therefore infer at once the convergence, in the vicinity of (0, 0), of the last power series in (20). For $m=1$ equation (22) becomes

$$s_{kl}^{(1)} = M\Big[\sum_{i=0}^{k-1} \bar{s}_{il}^{(1)} + \sum_{i=0}^{k} \sum_{j=0}^{l-1} \bar{s}_{ij}^{(1)} + \sum_{i=0}^{k} \sum_{j=0}^{l} \bar{s}_{ij}^{(0)} \Big].$$

Taking account of the obvious relations $\bar{s}_{ij}^{(0)} \leqq \bar{s}_{ij}^{(1)} \leqq \cdots \leqq \bar{s}_{ij}^{r}$, we have

$$\begin{aligned} \bar{s}_{kl}^{(1)} &= M \sum_{i=0}^{k-1} \bar{s}_{il}^{(1)} + M \sum_{i=0}^{k} \bar{s}_{il}^{(0)} + (M+1) \bar{s}_{k,l-1}^{(1)} \\ &= M \sum_{i=0}^{k-1} (\bar{s}_{il}^{(1)} + \bar{s}_{il}^{(0)}) + M\Big[M \sum_{i=0}^{k-1} \bar{s}_{il}^{(0)} + (M+1) \bar{s}_{k,l-1}^{(0)} \Big] + (M+1) \bar{s}_{k,l-1}^{(1)} \\ &< (M+1)^2 \Big[\sum_{i=0}^{k-1} \bar{s}_{il}^{(1)} + \bar{s}_{k,l-1}^{(1)} \Big] = S_{kl}^{(1)}. \end{aligned}$$

The $S_{kl}^{(1)}$, thus defined by a recurrence relation exactly like (16), are the coefficients of a power series convergent in the vicinity of (0, 0); we therefore infer the same property for the next to the last series in (20). It is now easy to establish by induction the inequality

$$s_{kl}^{(j)} < (M+1)^{j+1} \Big[\sum_{i=0}^{k-1} \bar{s}_{il}^{(j)} + \bar{s}_{k,l-1}^{(j)} \Big],$$

from which follows at once the convergence, in the vicinity of (0, 0), of all the remaining power series in (20).

If there correspond to (ρ_0, σ_0) series like (18) except for the replacement of t by τ, or series of the type (19), their convergence may be proved in like manner; in each of these cases the $g_{0s}(\rho_0, \sigma_0, k, l, m, p)$ (cf. (3)) play a rôle like that of $g_{s0}(\rho_0, \sigma_0, k, l, m, p)$ above. Hence we conclude

THEOREM II. *Corresponding to each irregular characteristic number-pair* (ρ_0, σ_0) *whose irregularity is due to the failure of assumption* (b) *alone, there exist a certain number of solutions of the equation* (1) *which are analytic in the vicinity of the place* (0, 0) *and are expressed there by the series exhibited in* §3.

The italicized additional statements following Theorem I *also apply here.*

To complete the discussion of this and the preceding section we observe that $g(\rho_0, \sigma_0, k, l, 0, 0)$ can vanish for at most a finite number of pairs of values k, l when assumption (b) alone fails. This follows immediately since (i) for $k \geqq K$ and $l \geqq L$, $g(\rho_0, \sigma_0, k, l, 0, 0)$ is not zero; and (ii) for any particular $k[l]$ less than $K[L]$, $g(\rho_0, \sigma_0, k, l, 0, 0)$ is a polynomial of degree n in $r^l[q^k]$ and so can vanish for only a finite number of values of $l[k]$. Therefore only series with a finite number of logarithmic terms can occur.

5. **Other irregular characteristic number-pairs.** It is clear that in certain cases of irregular characteristic number-pairs due to the failure of assumptions (a) or (c) the method employed by the author in II can be used to bring the equation (1) into a form to which the work of the preceding sections is applicable; these cases occur when the coefficients (2) are of the form

$$a_{ij}(x, y) = x^{\alpha_{ij}} y^{\beta_{ij}} \cdot [\text{convergent power series in } x \text{ and } y] \quad (i, j = 0, 1, \cdots, n),$$

the α_{ij} and β_{ij} being integers $\geqq 0$.

Let $a_i (i = 0, 1, \cdots, n)$ denote the least of the $n+1$ exponents $\alpha_{ij} (j = 0, 1, \cdots, n)$; similarly let $b_j (j = 0, 1, \cdots, n)$ represent the least of the exponents $\beta_{ij} (i = 0, 1, \cdots, n)$. First consider the case of $b_0 = b_n = 0$. We plot the points a_i in the ia-plane and draw a broken line above or upon which all these points lie and of which each vertex and end point is a point of the set. Then to a segment of slope μ' there corresponds in general an appropriate number of formal solutions like (6) or those of §3 except for the additional factor

$$q^{(\mu'/2)(t^2-t)}. \tag{23}$$

Secondly, in the case of $a_0 = a_n = 0$ the situation is entirely parallel, the additional factor in the solution being

$$r^{(\nu'/2)(r^2-r)}. \tag{24}$$

If neither of these conditions, $b_0 = b_n = 0$ and $a_0 = a_n = 0$, is satisfied, a combination of the two processes here indicated may be used, the solutions obtained containing factors of both types (23) and (24). The solutions can be shown to converge in the vicinity of (0, 0) if after subjecting (1) to the transformation

$$f(x,y) = q^{(\mu'/2)(t^2-t)}r^{(\nu'/2)(\tau^2-\tau)}g(x,y),$$

the resulting equation has a characteristic equation whose solutions satisfy assumptions (a), (c), (d), and (e).

It is quite clear that assumptions (c), (d), and (e) are essential to the convergence proofs only, not to the existence of formal solutions. In this connection we may point out that the present work sheds a certain amount of light upon the cases which gave rise to the question proposed at the close of §2 of I; under the definitions set forth here we have in these cases an irregularity, due to the failure of assumption (c).

Reference should also be made in this section to the method of treating irregularities which has been elaborated in §3 of I and which is sometimes efficacious.

We shall not attempt here to answer the question as to what is the most general solution of the equation (1). It may be remarked, however, that a linear combination of any finite number of particular solutions, in which the coefficients are arbitrary functions $p_i(x, y)$ satisfying the conditions

$$p_i(qx,y) = p_i(x,ry) = p_i(x,y), \tag{25}$$

is also a solution.

6. **The non-homogeneous equation.** In this section we consider briefly the equation obtained by adding to the left member of (1) a known function $b(x, y)$ of the same character as the $a_{ij}(x, y)$:

$$\sum_{i,j=0}^{n} a_{ij}(x,y)f(q^ix,r^jy) + b(x,y) = 0. \tag{26}$$

For this equation we set forth

THEOREM III. *If among the number-pairs $(q^i, r^j)(i, j=0, 1, \cdots)$ there are exactly p_1 which satisfy the equation* (3), *each being counted according to its multiplicity in ρ, the equation* (26) *has a formal solution*

$$(\text{P.S.}) + t(\text{P.S.}) + \cdots + t^{p_1}(\text{P.S.}),$$

in which the coefficients of the power series may be calculated by substitution in (26); *if exactly p_2 of these number-pairs satisfy* (3) *when each is counted according to its multiplicity in σ, there is a second formal solution*

$$(\text{P.S.}) + \tau(\text{P.S.}) + \cdots + \tau^{p_2}(\text{P.S.});$$

there may also be other formal solutions involving t and τ simultaneously. If assumption (c) *is satisfied and if we have also*

$$\sum_{i=0}^{n} a_{in00}q^{ki} \neq 0 \qquad (k = 0,1, \cdots, K' - 1),$$

$$\sum_{i=0}^{n} a_{nj00}r^{lj} \neq 0 \qquad (l = 0,1, \cdots, L' - 1),$$

where K' and L' are determined in the same manner as were K and L, the methods of §§2, 4 *may be employed to prove that each formal solution converges in the vicinity of* (0, 0) *and so represents an analytic solution there.*

It is clear that the sum of a particular solution of (26) and a linear combination of any finite number of solutions of the associated homogeneous equation (1), with arbitrary coefficients $p_i(x, y)$ satisfying the condition (25), is also a solution of (26).

BROWN UNIVERSITY,
PROVIDENCE, R. I.

LES FONCTIONS POLYGÈNES COMME INTÉGRALES D'ÉQUATIONS DIFFÉRENTIELLES*

PAR
G. CALUGARÉANO

Les fonctions polygènes. En 1912 M. Pompeiu† a abordé pour la première fois d'une façon directe l'étude des fonctions continues d'une variable complexe z en renonçant aux classiques conditions de Cauchy, qui ont l'inconvénient de trop particulariser la notion de fonction d'une variable complexe.

Récemment, M. E. Kasner a employé le terme *polygène* pour marquer cette opposition nette entre le nouveau et l'ancien point de vue. Cette dénomination se justifie aussi par ce fait que, pendant qu'une fonction monogène de z possède *une seule* dérivée en chaque point régulier, une fonction polygène en possède plusieurs et même une infinité. Ainsi, M. Kasner a reconnu‡ que, étant donnée

$$f(z) = P(x, y) + iQ(x, y) \tag{1}$$

telle que P et Q possèdent des dérivées partielles du premier ordre, la dérivée de $f(z)$ suivant la direction faisant un angle ϕ avec OX est représentée par la formule suivante, les notations que nous employons étant légèrement différentes de celles déjà employées,

$$\frac{df}{dz_\phi} = \delta[f] + e^{-2i\phi}\cdot\rho[f], \tag{2}$$

avec

$$\delta[f] = \frac{1}{2}\left[\frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + i\left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right)\right],$$
$$\rho[f] = \frac{1}{2}\left[\frac{\partial P}{\partial x} - \frac{\partial Q}{\partial y} + i\left(\frac{\partial Q}{\partial x} + \frac{\partial P}{\partial y}\right)\right]. \tag{3}$$

Mais tous ces calculs se simplifient beaucoup en prenant comme nouvelles variables

* Presented to the Society, February 23, 1929; received by the editors in December, 1928.

† Rendiconti del Circolo Matematico di Palermo, vol. 33 (1912, 1st semester), p. 112; vol. 35 (1913, 1st semester), p. 277.

‡ Science, vol. 66 (1927), p. 581; Proceedings of the National Academy of Sciences, vol. 13 (1928), p. 75; Bulletin of the American Mathematical Society, vol. 34 (1928), p. 495 and p. 561; these Transactions, vol. 30 (1928), p. 803.

$$z = x + iy\,; \quad \bar{z} = x - iy$$

comme M. Al. Proca l'a remarqué d'abord.* On a

$$x = (z + \bar{z})/2\,; \quad y = (z - \bar{z})/(2i)\,. \tag{4}$$

Toute fonction polygène devient alors une fonction de z par l'intermédiaire de z et $\bar{z}$, et l'on a

$$\frac{df}{dz_\phi} = \frac{\partial f}{\partial z} + \frac{\partial f}{\partial \bar{z}}\frac{d\bar{z}}{dz} = \frac{\partial f}{\partial z} + e^{-2i\phi}\cdot\frac{\partial f}{\partial \bar{z}}, \tag{5}$$

ce qui montre que les deux paramètres différentiels du premier ordre que constituent δ et ρ se réduisent à des dérivées partielles

$$\delta[f] = \frac{\partial f}{\partial z}\,; \quad \rho[f] = \frac{\partial f}{\partial \bar{z}}. \tag{6}$$

$\rho[f]$ a été défini d'abord par M. Pompeiu qui l'a appelé *dérivée aréolaire* en vertu de la relation

$$2i\cdot\rho[f] = \lim \frac{1}{(\sigma)}\int_\sigma f(z)dz, \tag{7}$$

où σ représente un contour fermé entourant le point z_0 considéré et (σ) désigne l'aire que ce contour définit. La limite est prise en faisant tendre l'aire et la longeur de σ vers zéro de façon que z_0 reste toujours intérieur à σ. M. Kasner a proposé d'appeler $\delta[f]$ *dérivée moyenne*. Nous employerons aussi ces deux termes. Il est clair que si P et Q sont dérivables plus d'une fois, la formule (5) est susceptible d'être réitérée. On aura† en général*

$$\frac{d^n f}{dz_\phi^n} = \left(\frac{\partial}{\partial z} + e^{-2i\phi}\cdot\frac{\partial}{\partial \bar{z}}\right)^n f. \tag{8}$$

La notion d'intégrale polygène d'une équation différentielle. La notion de dérivée d'une fonction de z étant ainsi élargie, on se pose tout naturellement des questions telles que la détermination d'une fonction polygène dont la dérivée est donnée, et plus généralement on est amené à se demander si une équation différentielle

$$F[z, y, y', y'', \cdots, y^{(n)}] = 0 \tag{9}$$

dans laquelle on suppose avoir remplacé les dérivées par les expressions (8), est susceptible d'être vérifiée par une fonction polygène $y(z)$, *quelque soit la*

* Note added in proof by E. Kasner: The higher derivatives employed by M. Calugaréano are those I call rectilinear, a special case of the general curvilinear derivatives defined in my papers, where minimal coördinates are also employed.

† On trouvera un résumé de ces résultats dans notre note des Comptes Rendus, vol. 186 (1928), p. 930.

direction ϕ qui entre dans (8). Nous verrons par la suite que la réponse est affirmative.

Fonctions polygènes analytiques. Nous nous limiterons ici à chercher les intégrales polygènes analytiques en z, et nous appelerons ainsi les fonctions polygènes de z admettant en tout point régulier de leur domaine d'existence un développement en série double de puissances de z et $\bar{z}$. Un élément tel que

$$f(z) = \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} a_{m,n} z^m \bar{z}^n \tag{10}$$

convergera *au moins* à l'intérieur d'un cercle de rayon ρ_1 tel que

$$\frac{1}{\rho_1} = \limsup_{m+n\to\infty} (|a_{m,n}|)^{1/(m+n)}$$

de quelque façon que $(m+n)$ tende vers l'infini. Dès lors, le prolongement *à la Weierstrass* nous permettra de définir $f(z)$ dans tout son domaine d'existence, à moins que cette fonction ne possède des *lignes singulières fermées*. Dans ce cas le prolongement à la Weierstrass peut devenir insuffisant pour définir $f(z)$; tel est le cas de $1/(1-z\bar{z})$ dont le développement autour de l'origine ne saurait être prolongé au delà du cercle unité, car $1-z\bar{z}=1-r^2$ s'annulle tout le long de ce cercle; pourtant $1/(1-z\bar{z})$ existe dans tout le plan et son *expression analytique* permet de suppléer à l'insuffisance du prolongement.

Les fonctions polygènes analytiques étant définies par la possibilité d'un développement tel que (10) en tout point régulier, remarquons que ces fonctions sont *indéfiniment* dérivables. A moins qu'elles ne possèdent des coupures fermées, elles sont des fonctions quasi analytiques de z, et ceci veut dire justement qu'elles sont déterminées dans tout leur domaine d'existence une fois qu'elles sont connues dans un domaine d'aire non nulle, aussi petit qu'il soit.

Théorème établissant une liaison entre intégrales polygènes et intégrales monogènes. Nous nous occuperons donc uniquement des intégrales polygènes analytiques d'équations telles que (9). Soit donc $y(z)$ une fonction polygène analytique satisfaisant à (9) quelque soit ϕ, et considérons un segment AB parallèle à OX et entièrement intérieur au domaine de régularité de $y(z)$. Soit d sa distance à OX. On a

$$y(z) = G(z, \bar{z}),$$

G étant une fonction analytique des deux lettres qu'elle contient. Le long de AB, on aura

$$y = G(x + id,\ x - id).$$

Comme $y(z)$ est régulière sur AB, cette fonction de x est analytique au sens ordinaire. Il s'en suit, par prolongement, que

$$Y(z) = G(z,\ z - 2id)$$

est une *intégrale monogène* de l'équation (9). On voit donc que toute intégrale polygène analytique de (9) nous fait connaître une intégrale monogène par la simple substitution de $z-2id$ à $\bar{z}$. Et ceci introduit une constante d en plus, qui est en général arbitraire.

Les équations admettant des intégrales polygènes forment une classe particulière. Il est facile de voir qu'une équation différentielle du premier ordre ne saurait posséder des intégrales polygènes. En effet, on devrait avoir

$$F\left(z,\ y,\ \frac{\partial y}{\partial z} + e^{-2i\phi}\cdot\frac{\partial y}{\partial \bar{z}}\right) = 0$$

quelque soit ϕ. Ceci exige

$$\frac{\partial y}{\partial \bar{z}} = 0,$$

donc y est monogène en z.

Les équations qui nous intéressent sont donc au moins d'ordre 2. J'ai traité ailleurs* et d'une façon complète, le cas général de l'équation du second ordre. On trouve que la plus générale équation du second ordre, possédant des intégrales polygènes, est

$$(11)\quad y'' + \frac{\partial \omega}{\partial y}y'^2 + 2\frac{\partial \omega}{\partial z}y' + e^{-\omega}\left\{\Phi + \int\left[\left(\frac{\partial \omega}{\partial z}\right)^2 + \frac{\partial^2 \omega}{\partial z^2}\right]e^{\omega}dy\right\} = 0,$$

où $\omega(z, y)$ est une fonction *arbitraire* monogène en z et y, $\Phi(z)$ une fonction monogène *arbitraire* de z. Les intégrales polygènes de cette équation s'obtiennent en résolvant

$$\alpha\bar{z} + \beta z + \gamma = \mu(z, y) = \int_0^y e^{\omega(z,s)}ds$$

par rapport à y. On voit que l'intégrale polygène dépend de *trois* constantes arbitraires α, β, γ, donc elle est *plus générale* que l'intégrale monogène; l'intégrale monogène Y s'obtient, selon le théorème déjà indiqué, en changeant $\bar{z}$ en $z-C$. On trouve ainsi $Y(z)$ en résolvant

$$az + b = \mu(z, Y)$$

par rapport à Y.

Donc, l'intégration au sens classique est complète pour cette équation et il suffit d'une quadrature. Les calculs se compliquent notablement quand on essaye de traiter des équations d'ordre supérieur. La condition qui caractérise la classe d'équations différentielles possédant des intégrales polygènes paraît être assez cachée, peut être assez compliquée même.

* Voir ma note déja citée et surtout la suite, Comptes Rendus, vol. 186, p. 1406.

Mais, si toute équation différentielle ne possède pas des intégrales polygènes on voit facilement que toute fonction polygène, contenant ou non des constantes arbitraires, est intégrale d'une certaine équation différentielle. Il suffira de considérer

$$y(z) = \Phi(z, \bar{z}, c),$$

d'où

$$y' = \frac{\partial \Phi}{\partial z} + \mu \frac{\partial \Phi}{\partial \bar{z}},$$

$$y'' = \frac{\partial^2 \Phi}{\partial z^2} + 2\mu \frac{\partial^2 \Phi}{\partial z \partial \bar{z}} + \mu^2 \frac{\partial^2 \Phi}{\partial \bar{z}^2},$$

$$y''' = \frac{\partial^3 \Phi}{\partial z^3} + 3\mu \frac{\partial^3 \Phi}{\partial z^2 \partial \bar{z}} + 3\mu^2 \frac{\partial^3 \Phi}{\partial z \partial \bar{z}^2} + \mu^3 \frac{\partial^3 \Phi}{\partial \bar{z}^3},$$

avec

$$\mu = e^{-2i\phi}.$$

En éliminant C, $\bar{z}$, μ entre ces quatre équations, on aura une équation différentielle que $y(z)$ vérifie.

Signalons ici certaines classes d'équations dont l'intégration peut être particulièrement facile vis-à-vis du cas général.

Rôle des équations aux dérivées aréolaires. Considérons des équations dont le premier membre est une fonction algébrique et entière par rapport aux dérivées $y', y'', \cdots, y^{(n)}$ mais pouvant être une fonction monogène transcendente de z et y.

Prenons donc

$$y^{(n)} - P[z, y ; y', y'', \ldots, y^{(n-1)}] = 0, \tag{12}$$

et soit

$$A(z, y)[y']^{q_1}[y'']^{q_2} \ldots [y^{(n-1)}]^{q_{n-1}}$$

un terme homogène du polynôme P. Appelons *poids* de ce terme la quantité

$$p = q_1 + 2q_2 + 3q_3 + \ldots + (n-1)q_{n-1},$$

et *poids du polynôme* le plus grand poids réalisé par ses termes. Limitons-nous aux équations d'ordre n et de poids $\leqq n$. Par cette particularisation nous ne faisons que renouveler les circonstances qui se sont présentées pour l'équation (11) dont l'intégration a pu être complètement achevée.

Pour intégrer (12) nous remplacerons les dérivées de $y(z)$ par leurs expressions (8), en obtenant ainsi au premier membre de (12) un polynôme de degré n en $\mu = e^{-2i\phi}$. L'égalité devant avoir lieu pour toute valeur de ϕ, on devra annuler les $(n+1)$ coefficients de ce polynôme en μ, ce qui ramène la recherche des intégrales polygènes à l'intégration d'un systéme (S_n) de

$(n+1)$ équations aux dérivées partielles en y. Les variables indépendantes seront z et $\bar{z}$ et le système sera d'ordre n.

Envisageons plus attentivement l'équation qu'on obtient en annullant le coefficient de μ^n. Il est facile de remarquer que ce coefficient, étant formé avec les coefficients des plus hautes puissances de μ dans chaque dérivée, sera un polynôme en $\partial y/\partial\bar{z}$, $\partial^2 y/\partial\bar{z}^2$, $\cdots$, $\partial^n y/\partial\bar{z}^n$ et ne contiendra *aucune* dérivée mixte telle que $\partial^{m+p}y/\partial z^m\partial\bar{z}^p$. C'est pourquoi nous appellerons *équation aux dérivées aréolaires* cette équation du système (S_n). Remarquons que pour *toute* équation du type (12) le système (S_n) contient au moins une équation aux dérivées aréolaires obtenue en annullant le coefficient de μ^n. Mais il se peut qu'il en existe plusieurs, pour des cas moins généraux. Or, l'existence de cette équation aux dérivées aréolaires simplifie l'intégration du système (S_n). Pour faire voir ceci considérons l'équation aux dérivées aréolaires correspondant à (11). Elle est de la forme

$$\frac{\partial^2 y}{\partial\bar{z}^2} + A(z,y)\left(\frac{\partial y}{\partial\bar{z}}\right)^2 = 0,$$

et s'intègre facilement par quadratures,

$$\frac{\dfrac{\partial^2 y}{\partial\bar{z}^2}}{\dfrac{\partial y}{\partial\bar{z}}} = -A(z,y)\frac{\partial y}{\partial\bar{z}},$$

$$\frac{\partial y}{\partial\bar{z}} = \Phi_1(z)e^{-\int_0^y A(z,s)ds},$$

$$\bar{z}\Phi_1(z) + \Phi_2(z) = \int_0^y e^{\int_0^t A(z,s)ds}dt = \mu(z,y),$$

où les fonctions monogènes arbitraires Φ_1 et Φ_2 jouent, pendant l'intégration, le rôle de deux constantes arbitraires.

Il en est de même pour toute équation aux dérivées aréolaires et ceci simplifie toujours l'intégration de (S_n). M. Pompeiu a été le premier à considérer une équation aux dérivées aréolaires en se demandant s'il existe des fonctions polygènes égales à leur dérivée aréolaire. Avec la définition (7) la question semble moins facile à priori. Mais elle revient à

$$\frac{\partial E}{\partial\bar{z}} = E(z,\bar{z}),$$

d'où

$$\Phi(z) + \bar{z} = \log(E); \qquad E = \Phi_1(z)\cdot e^{\bar{z}},$$

où Φ_1 et Φ désignent des fonctions monogènes. Il arrive de plus que toute équation aux dérivées aréolaires, meme si elle ne peut pas être intégrée complètement, est réductible à un ordre moindre; son ordre peut toujours etre réduit d'une unité.

En effet, soit

$$(13)\quad \frac{\partial^n y}{\partial \bar{z}^n} = \sum_{s=0}^{n} A_{p_1,q_1;\,p_2,q_2;\,\ldots;\,p_s,q_s}(z,\,y)\left[\frac{\partial^{p_1} y}{\partial \bar{z}^{p_1}}\right]^{q_1}\left[\frac{\partial^{p_2} y}{\partial \bar{z}^{p_2}}\right]^{q_2}\cdots\left[\frac{\partial^{p_s} y}{\partial \bar{z}^{p_s}}\right],$$

l'équation en question, les $A(z, y)$ étant monogènes en z et y en vertu des hypothèses faites sur (12). Pour intégrer cette équation, on considérera z comme un paramètre et l'on remarquera que $\bar{z}$, la véritable variable indépendante, ne figure pas explicitement dans cette équation. On prendra donc $\bar{z}$ comme fonction et y comme nouvelle variable indépendante, et l'ordre sera ainsi rabaissé d'une unité. Si la nouvelle équation peut être complètement intégrée on aura

$$(14)\qquad \bar{z} = F(z,\,y)$$

qui, résolue par rapport à y, donnera l'intégrale de (13).

Les n constantes arbitraires $\Phi_k(z)$ qui s'introduisent ainsi doivent être considérées comme fonctions arbitraires, monogènes, de z. Ceci nous précise la forme analytique de l'intégrale polygène comme fonction* de $\bar{z}$. Pour avoir son expression complète et pour s'assurer de son existence effective qui n'est pas encore établie, il reste à discuter les n autres équations de (S_n). Ceci peut se faire en substituant y donné par (14) dans ces n équations, ce qui permet d'en tirer des conditions supplémentaires pour les n fonctions arbitraires $\Phi_k(z)$ et pour les coefficients $A(z, y)$ de (12). Mais ces dernières conditions, qui précisent la classe d'équations (12) admettant des intégrales polygènes, peuvent s'obtenir par un calcul plus symétrique, quoique assez laborieux, en éliminant les dérivées de y entre les $(n+1)$ équations de (S_n).

Enfin, la substitution de $z-C$ à $\bar{z}$ transformera les intégrales polygènes ainsi obtenues en des intégrales monogènes de la même équation. Il n'est pas sûr à priori que celles-ci contiennent effectivement n constantes arbitraires.

On voit donc que l'étude des intégrales polygènes d'une équation différentielle est de nature à nous fournir des renseignements sur les intégrales monogènes de la même équation, que le théorème classique de Cauchy concerne uniquement.

* On remarque facilement que, pour des équations du type (12), celle-ci se réduit à un polynome de degré $(n-1)$ en $\bar{z}$, à coefficients monogènes en z, toutes les fois que le poids de P est $<n$. Il en est tout autrement quand il est $=n$.

UNIVERSITY OF PARIS,
PARIS, FRANCE

THE FOUNDATIONS OF THE CALCULUS OF VARIATIONS IN THE LARGE IN m-SPACE (FIRST PAPER)*

BY

MARSTON MORSE

1. **Introduction.** The calculus of variations in the large as developed by the author may be roughly divided into two parts.

The first part involves the characterization of the number of conjugate points or focal points on a given extremal in terms of the type number of a basic quadratic form, and furnishes a connecting link between the theory of such extremals and the theory of critical points of functions.

The second part goes beyond the properties of one extremal alone; it involves families of extremals, families both closed and open. It necessarily leads to considerations which belong to analysis situs. It seeks to prove the existence of extremals on open or closed manifolds regardless of whether or not the extremals give minima.

The present paper consists of three parts of which the first two set up and study the fundamental forms associated respectively with an extremal joining two points, and an extremal cut transversally by a given manifold.

The third part presents a relatively complete picture of the theory in the large of extremals from a given point cut transversally by a given closed manifold. It gives strong existence theorems immediately applicable, for example, to the question of how many straight lines one can draw from a point normal to a given manifold. For the case where the manifold is homeomorphic to an $(m-1)$-sphere it can be shown that the relations found in this part are exhaustive.

The first two parts taken with the *deformation theory* previously developed by the author† will culminate in a second paper which will give the central existence theorems in the large for extremals joining two fixed points.

The first two parts taken in quite another way will lead to a generalization of the famous Sturm separation and comparison theorems for the case of a system of n linear, second-order, homogeneous, self-adjoint differential

* Presented to the Society as part of an address given at the request of the program committee April 6, 1928, under the title *The critical points of functions and the calculus of variations in the large*; see also Bulletin of the American Mathematical Society, vol. 35 (1929), pp. 38–54.

Received by the editors in April, 1929.

† Marston Morse, *The foundations of a theory in the calculus of variations in the large*, these Transactions, vol. 30 (1928), pp. 213–274.

BOSTON UNIVERSITY
COLLEGE OF LIBERAL ARTS
LIBRARY

equations. The methods and the results of this paper are also extensible to the case of two variable end points.

I. Extremals through fixed points

2. **The integrand*** $F(x, r)$. Let R be a closed region in the space of the variables $(x_1, \cdots, x_m) = (x)$. Let

$$F(x_1, \cdots, x_m, r_1, \cdots, r_m) = F(x, r)$$

be a function of class C''' for (x) in R and (r) any set not (0). We suppose further that F is homogeneous in that

$$F(x_1, \cdots, x_m, kr_1, \cdots, kr_m) = kF(x, r) \tag{2.1}$$

for any positive constant k. We shall start with the integral in the usual parametric form

$$J = \int_{t_0}^{t_1} F(x, \dot{x})dt \tag{2.2}$$

where $(\dot{x})$ stands for the set of derivatives of (x) with respect to t.

Let g be an extremal segment lying in R. We suppose that g is an ordinary curve of class C'. Denote the second partial derivative of $F(x, r)$ with respect to r_i and r_j by F_{ij}. Along g we assume that F is positively regular, that is, that†

$$F_{ij}(x, \dot{x})\eta_i\eta_j > 0 \qquad (i, j = 1, 2, \cdots, m) \tag{2.3}$$

for $(x, \dot{x})$ taken along g, and for (η) any set not (0) nor proportional to $(\dot{x})$.

Because of the homogeneity relation (2.1) we have the m identities

$$r_i F_{ij}(x, r) \equiv 0 \qquad (j = 1, 2, \cdots, m) \tag{2.4}$$

from which it appears that

$$|F_{ij}| \equiv 0. \tag{2.5}$$

---

*For that which concerns the classical theory of the calculus of variations in space the reader may refer to the following articles:

Mason and Bliss, *The properties of curves in space which minimize a definite integral*, these Transactions, vol. 9 (1908), pp. 440–466.

Bliss, *The Weierstrass E-function for problems of the calculus of variations in space*, these Transactions, vol. 15 (1914), pp. 369–378.

Bliss, *Jacobi's condition for problems of the calculus of variations in parametric form*, these Transactions, vol. 17 (1916), pp. 195–206.

Hadamard, *Leçons sur le Calcul des Variations*, vol. 1, Paris, 1910.

Carathéodory, *Die Methode der geodätischen Äquidistanten und das Problem von Lagrange*, Acta Mathematica, vol. 47 (1926), pp. 199–236.

†Here and elsewhere we adopt the convention that the repetition of a subscript indicates a summation with respect to that subscript.

The generalized Weierstrassian function $F_1(x, r)$ can here be conveniently defined as the sum of the principal minors of order $m-1$ of the determinant (2.5) divided by the sum of the squares of the r_i's.*

For $(x, \dot{x})$ on g we now form the characteristic determinant $D(\lambda)$ of the determinant (2.5). We see that one of the roots of $D(\lambda)=0$ must be zero on account of (2.5). The remaining roots must be positive on account of (2.3). A consequence is that the coefficient of λ in $D(\lambda)$ must be negative. But this coefficient is

$$-(\dot{x}_1^2 + \cdots + \dot{x}_m^2)F_1(x, \dot{x})$$

so that F_1 is thereby proved to be positive along g.

3. **The invariance of our hypotheses under admissible transformations.** By an admissible transformation from the variables (x) to a new set of variables (z) will be understood a transformation of the form

$$x_i = \phi_i(z_1, \cdots, z_m) \qquad (i = 1, 2, \cdots, m), \tag{3.1}$$

where the functions ϕ_i are of class C''' in their arguments and where the jacobian of the ϕ_i's with respect to the z_i's is never zero. We obtain a new integrand $G(z, r)$ by setting

$$G(z_1, \cdots, z_m, r_1, \cdots, r_m) = F\left(\phi_1, \cdots, \phi_m, \frac{\partial \phi_1}{\partial z_i} r_i, \cdots, \frac{\partial \phi_m}{\partial z_i} r_i\right). \tag{3.2}$$

The positive regularity of F along g implies the positive regularity of G along the transform of g, say γ. That is, it follows from (2.3) and (3.2) that

$$G_{ij}(z, \dot{z})\zeta_i\zeta_j > 0 \qquad (i, j = 1, 2, \cdots, m) \tag{3.3}$$

for $(z, \dot{z})$ on γ, and for (ζ) any set not (0) nor proportional to $(\dot{z})$. The proof of this statement appears readily if one obtains the second partial derivatives of G in terms of those of F.

That G possesses the same homogeneity property as does F, follows from (3.2).

4. **A transformation carrying g into a straight line.** The following lemma will greatly simplify our future work.

LEMMA. *There exists an admissible transformation T of the variables (x) into variables (z) under which the image of g is a segment of the z_1 axis.*

Suppose g is given in terms of its arc length in the form

$$x_i = h_i(s) \qquad (i = 1, 2, \cdots, m). \tag{4.1}$$

* See Bliss, second paper cited, p. 378, for a differently phrased but equivalent definition.

We now state the following preliminary proposition.

(A). Let the direction cosines of the direction tangent to g be represented by a point P_1 on a unit $(m-1)$-sphere S with center at the origin O. As P_1 moves continuously on S there can be chosen $m-1$ other points $P_2, \cdots, P_m$ on S which vary continuously with s, and are such that the points

$$O, P_1, \cdots, P_m$$

for no value of s lie on the same $(m-1)$-plane.

The proof of this proposition may be given in so many ways that it can very well be left to the reader.

Granting the truth of (A) consider the determinant b whose ith column gives the coördinates of the point P_i. We now replace the elements of the last $m-1$ columns of b by analytic approximations taken so close that the new determinant $a(s)$ is never zero. Let $a_{ij}(s)$ be the ijth element of $a(s)$. Our transformation T can now be defined as follows:

$$x_i = h_i(z_1) + a_{ik}(z_1)z_k \qquad (k = 2,3, \cdots, m\,;\ i = 1,2, \cdots, m) \tag{4.2}$$

where z_k is taken in the neighborhood of $z_k = 0$, and z_1 may take on the same values as s. We have for $z_k = 0$

$$\frac{D(x_1, \cdots, x_m)}{D(z_1, \cdots, z_m)} = a(z_1) \neq 0, \tag{4.3}$$

so that the lemma follows readily.

5. **The integral in non-parametric form.** According to the result of the last section there exists an admissible transformation T under which g is carried into a segment γ of the z_1 axis. We suppose that the positive sense along g corresponds to the positive sense of the z_1 axis. Suppose that under T the new integrand G is given by (3. 2). We now set

$$(z_1, \cdots, z_m) = (x, y_1, \cdots, y_n) = (x, y), \qquad n = m - 1,$$

and further set

$$\begin{aligned} f(x,y,p) &= f(x, y_1, \cdots, y_n, p_1, \cdots, p_n) \\ &= G(x, y_1, \cdots, y_n, 1, p_1, \cdots, p_n) \end{aligned} \tag{5.1}$$

for (x, y) neighboring γ and for any set (p).

For ordinary curves neighboring γ for which $\dot{z}_1 > 0$, our integral J now takes the non-parametric form

$$J = \int_a^b f(x, y, y')dx. \tag{5.2}$$

It will be convenient to set

$$f^0_{p_ip_j} = A_{ij}(x), \quad f^0_{p_iv_j} = B_{ij}(x), \quad f^0_{v_iv_j} = C_{ij}(x),$$

where the partial derivatives involved are evaluated for $(x, y, p) = (x, 0, 0)$, as indicated by the superscript zero.

We shall now reduce the regularity condition (3.3) *to the condition that*

$$A_{ij}(x)\zeta_i\zeta_j > 0 \qquad (i,j = 1, \cdots, n) \tag{5.3}$$

for every set $(\zeta) \neq (0)$.

Because G is homogeneous in (r) we have

$$G_{ij}(z,\dot{z})\dot{z}_j = 0. \tag{5.4}$$

In particular along γ we have

$$\dot{z}_2 = \cdots = \dot{z}_m = 0,$$

so that (5.4) reduces along γ to

$$G_{i1}(z,\dot{z})\dot{z}_1 = 0. \tag{5.5}$$

Since $\dot{z}_1 \neq 0$ along γ we infer from (5.5) that the elements in the first row and column of the matrix of the form (3.3) are zero along γ. With this understood the condition (3.3) taken along γ reduces with the aid of (5.1) to (5.3).

A particular consequence of (5.3) is that

$$|A_{ij}| \neq 0 \tag{5.6}$$

along γ.

6. **The invariance of the order of a conjugate point.** Let O be a point of g and (λ) the set of the m direction cosines of the direction of g at O. Let (λ) represent a point P on a unit $(m-1)$-sphere S with center at the origin. Let (α) be a set of parameters in an admissible* representation of S neighboring P, or, in other words, of the directions neighboring (λ). Suppose that $(\alpha) = (\alpha^0)$ corresponds to P. Let s be the arc length on the extremals issuing from O, measuring s from O in the positive senses of the extremals, and confining s to the set of values it has on g. It is well known that the extremals that issue from O with the directions determined by (α) can be represented in the form

$$x_i = x_i(s,\alpha) \qquad (i = 1,2, \cdots, m), \tag{6.1}$$

* That is, a representation in which the coördinates (x) are to be functions of class C''' of $m-1$ parameters (α) of such sort that not all of the jacobians of $m-1$ of the coördinates with respect to the parameters (α) are zero.

where the functions (6.1) and their first partial derivatives as to s are of class C''.

The jacobian of the functions (6.1), namely

$$\Delta(s) = \frac{D(x_1, \cdots, x_m)}{D(s, \alpha_1, \cdots, \alpha_n)}, \quad (\alpha) = (\alpha^0), \quad n = m - 1, \tag{6.2}$$

is clearly zero for $s=0$. That the zero of $\Delta(s)$ at $s=0$ is isolated can also be readily proved.* We come now to an important definition.

The points on g, not O, at which $\Delta(s)$ vanishes are called the conjugate points of O, and the order of the vanishing of $\Delta(s)$ at each of these points the order of that conjugate point.

We turn next to the non-parametric integral (5.2), and suppose the J. D. E.† set up in the non-parametric form corresponding to the extremal γ. Suppose that under the transformation T the point O on g corresponds to the point $x=a$ on γ. Let $D(x)$ be an n-square determinant whose n columns give respectively n linearly independent solutions of the J. D. E. which vanish at $x=a$. Any other such determinant will be a non-vanishing constant times $D(x)$. We shall first prove the following

LEMMA. *Under the transformation T the zeros of $\Delta(s)$ on g correspond to the zeros of $D(x)$ on γ, and the orders of corresponding zeros are the same.*

By virtue of the transformation T the extremals given by (6.1) can be represented in the space of the new variables (x, y) in the form

$$x = h(s, \alpha), \tag{6.3}$$

$$y_i = k_i(s, \alpha) \qquad (i = 1, 2, \cdots, n = m - 1), \tag{6.4}$$

where it appears that the jacobian

$$\Delta_1(s) = \frac{D(h, k_1, \cdots, k_n)}{D(s, \alpha_1, \cdots, \alpha_n)}, \qquad (\alpha) = (\alpha^0), \tag{6.5}$$

equals the jacobian $\Delta(s)$ divided by the non-vanishing jacobian of the transformation T. Thus $\Delta(s)$ and $\Delta_1(s)$ vanish together and to the same orders.

Now γ coincides with the x axis so that $h_s(s, \alpha^0) \neq 0$ on γ. We can thus solve (6.3) for s as a function of x and (α), substitute the result in (6.4), and so represent the extremals (6.1) in the form

* See Mason and Bliss, loc. cit., p. 447. It is easy to see that the determinant which multiplies U on p. 447 is not zero when (u, v) are a pair of parameters admissible in our sense.

† We write J.D.E. for the Jacobi differential equations.

$$y_i = w_i(x,\alpha) \qquad (i = 1,2,\cdots,n). \tag{6.6}$$

The jacobian

$$\Delta_2(x) = \frac{D(w_1,\cdots,w_n)}{D(\alpha_1,\cdots,\alpha_n)}, \qquad (\alpha) = (\alpha^0), \tag{6.7}$$

is related to the jacobian $\Delta_1(s)$ in that

$$h_s(s,\alpha^0)\Delta_2(x) \equiv \Delta_1(s), \tag{6.8}$$

where x and s correspond under (6.3). Hence $\Delta_2(x)$ and $\Delta(s)$ vanish at corresponding points and to the same orders.

Finally, as is well known, the jacobian $\Delta_2(x)$ consists of columns which satisfy the J. D. E. The elements of these columns all vanish at $x=a$, and lastly are linearly independent since $\Delta_2(x)\not\equiv 0$. Hence

$$\Delta_2(x) \equiv cD(x), \tag{6.9}$$

where c is a non-vanishing constant. The lemma follows directly.

The following statement is an easy corollary.

Under admissible space transformations conjugate points correspond to conjugate points and the orders of conjugate points are preserved.

7. **A lemma on the order of a conjugate point.** In terms of the determinant $D(x)$ of §6 we have the following lemma.

LEMMA. *If $D(x)$ vanishes to the rth order at $x=a$ the rank of $D(a)$ is $n-r$.*

For simplicity suppose $a=0$. Let $a_{ij}(x)$ be the general element of $D(x)$. Let $\boldsymbol{b}(x)$ be a matrix whose general element is

$$b_{ij}(x) = a_{ij}(0) + a_i'_j(0)x. \tag{7.1}$$

According to the theory of λ matrices (here $\lambda=x$) there exist n-square non-singular matrices $\mathbf{c}$ and $\boldsymbol{d}$ of constant elements such that*

$$\boldsymbol{cbd} = \mathbf{e}(x) \tag{7.2}$$

where $\mathbf{e}(x)$ is a matrix all of whose elements are identically zero except those in the principal diagonal.

With $\mathbf{c}$ and $\boldsymbol{d}$ so determined let us transform the dependent variables (η) of the J. D. E. into variables (w) by setting

$$w_i = c_{ij}\eta_j \qquad (i,j = 1,2,\cdots,n), \tag{7.3}$$

* See Bôcher, *Introduction to Higher Algebra*, Chapter XX.

where c_{ij} is the general element of $\mathbf{c}$. The transformed J. D. E. will have for solutions the columns of the matrix

$$\mathbf{ca}(x). \tag{7.4}$$

It will also have for solutions the columns of the matrix

$$\mathbf{ca}(x)\mathbf{d} = \mathbf{g}(x). \tag{7.5}$$

It follows from (7.1), (7.2), and (7.5), that

$$e_{ij}(x) = g_{ij}(0) + g_i'_j(0)x. \tag{7.6}$$

Further I say that none of the elements in the principal diagonal of $\mathbf{e}(x)$ are identically zero. For if so the corresponding column of $\mathbf{g}(x)$ would represent a solution of the transformed J. D. E. whose coördinates as well as their derivatives would all be zero at $x=0$, and so would be identically zero. Hence $|g(x)|$ would be identically zero contrary to (7.5).

Thus the elements in the principal diagonal of $\mathbf{e}(x)$ if suitably reordered must be of the form

$$a_1x, \cdots, a_kx, \quad a_{k+1}x + b_{k+1}, \cdots, a_nx + b_n$$

where k is an integer from 1 to n inclusive, and none of the constants

$$a_1, \cdots, a_k, b_{k+1}, \cdots, b_n$$

are zero.

By virtue of the relations (7.6) between $\mathbf{e}(x)$ and $\mathbf{g}(x)$ it appears then that $|g(x)|$ vanishes to the order k at $x=0$, and has the rank $n-k$ when $x=0$. From (7.5) it follows that $|g(x)|$ and $|a(x)|$ vanish to the same order at $x=0$, and have the same rank when $x=0$. Since $D(x)$ equals $|a(x)|$ it vanishes to the order k at $x=0$, and has the rank $n-k$ when $x=0$. Thus the lemma is proved.

8. **Two additional lemmas.** The preceding lemma leads to the following.

LEMMA 1. *If $x=b$ is a conjugate point of $x=a$ of the rth order there will be just r linearly independent solutions of the J. D. E. which vanish at both $x=b$ and $x=a$.*

Any solution which vanishes at a must be linearly dependent on the columns of $D(x)$, and conversely any such linear combination will vanish at a. If now we turn to b and write down the n conditions that a linear combination of the columns of $D(x)$ vanish at b we see from the preceding lemma that there are just r linearly independent combinations that satisfy the n conditions. The lemma follows directly.

We note the following consequence.

If $x=b$ is a conjugate point of $x=a$ of the rth order, then $x=a$ is a conjugate point of $x=b$ of the rth order.

For future use Lemma 1 will be generalized as follows. The preceding proofs hold for the generalization.

LEMMA 2. *Let $D(x)$ be a determinant whose n columns are n linearly independent solutions of the J. D. E. If $D(x)$ vanishes to the rth order at $x=c$ the number of linearly independent solutions dependent on the columns of $D(x)$ which vanish at $x=c$ equals r.*

9. **The fundamental quadratic form.** Our extremal g has been transformed into a segment γ of the x axis in the space of the variables

$$(x,y) = (x,y_1, \cdots, y_n).$$

We suppose γ bounded by the points $x=a$ and $x=b$ $(a<b)$.

Let us cut across the portion of γ for which $a<x<b$ with p successive n-planes t_i, perpendicular to γ, cutting γ respectively in points at which $x=x_i$. These n-planes divide γ into $p+1$ successive segments. Suppose these n-planes are placed so near together that no one of these $p+1$ segments contains a conjugate point of its initial end point. Let P_i be any point on t_i near γ. Let the points on γ at which $x=a$ and $x=b$, respectively, be denoted by A and B.

If the points

$$A, P_1, \cdots, P_p, B \tag{9.1}$$

are sufficiently near γ they can be successively joined by extremal segments neighboring γ. Let the resulting broken extremal be denoted by E. Let (u) be the set of $\mu=pn$ variables of which the first n are the coördinates (y) of P_1, the second those of P_2, and so on, until finally the last n are the coördinates (y) of P_p. The value of the integral J taken along E will be a function of the variables (u), and will be denoted by

$$J(u_1, \cdots, u_\mu) = J(u).$$

When $(u)=(0)$, E will become γ. Considerations similar to these taken up by the author* for the case $n=1$ show that $J(u)$ is of class C'' for (u) near (0). It clearly has a critical point when $(u)=(0)$. We come next to the terms of second order in $J(u)$, that is, to the fundamental form

$$\sum J_{ij}u_iu_j \qquad (i,j = 1,2, \cdots, \mu), \tag{9.2}$$

* See Morse, loc. cit., pp. 217–219.

where ${}_0J_{ij}$ is the partial derivative of J with respect to u_i and u_j, and is evaluated for $(u)=(0)$, as indicated by the subscript 0.

Our first problem is to determine the rank and signature of this form.

10. The fundamental form and the J. D. E. As is conventional we set

$$\Omega(x,\eta,\eta') = A_{ij}\eta_i'\eta_j' + 2B_{ij}\eta_i'\eta_j + C_{ij}\eta_i\eta_j \qquad (i,j=1,\cdots,n) \tag{10.1}$$

where the arguments (x, y, p) in the partial derivatives of $f(x, y, p)$ are $(x, 0, 0)$. The J. D. E. corresponding to the extremal γ are

$$\Omega_{\eta_i} - \frac{d\Omega_{\eta_i'}}{dx} = 0 \qquad (i=1,\cdots,n). \tag{10.2}$$

It will be convenient to suppose that the points (9.1) lie not only in the space (x, y) but also in the space (x, η) of the J. D. E. A curve representing a solution of the J. D. E. will be called a *secondary extremal.* We see that successive points of (9.1) can be joined by secondary extremals. The set of these secondary extremals will form a broken secondary extremal which we may regard as *determined* by the set (u) of §9. With this understood we can prove the following

LEMMA. *The fundamental form may be represented as follows:*

$${}_0J_{ij}u_iu_j = \int_a^b \Omega(x,\eta,\eta')dx \qquad (i,j=1,\cdots,\mu), \tag{10.3}$$

where (η) stands for the set of n functions of x which give the coördinates other than x of a point on the broken secondary extremal determined by (u).

To prove this let e be a variable neighboring $e=0$, and suppose the integral J evaluated along the broken extremal E *determined* when the set (u) is replaced by

$$(eu_1,\cdots,eu_\mu). \tag{10.4}$$

This value of J is given by the function

$$J(eu_1,\cdots,eu_\mu). \tag{10.5}$$

If now we differentiate the function (10.5) twice with respect to e, and then put $e=0$, we get the left hand member of (10.3).

If on the other hand we take the integral J along the curve E, differentiate under the integral sign twice with respect to e, and then set $e=0$, we will get the right hand member of (10.3). Here the functions η_i will be the partial derivatives with respect to e of the coördinates y_i of the point on E, and will thus represent a broken secondary extremal. If the conditions that E be the

broken extremal determined by (10.4) be differentiated with respect to e it will follow that the functions (η) in (10.3) give that particular broken secondary extremal which is determined by (u).

The identity (10.3) follows at once.

11. The rank of the fundamental form. The theorem here is the following.

THEOREM 1. *If $x=b$ is a conjugate point of $x=a$ of the rth order the rank of the matrix of the form*

$$(11.1) \qquad {}_0J_{ij}u_iu_j = Q \qquad (i,j = 1,2,\cdots,\mu)$$

is $\mu-r$. If $x=b$ is not conjugate to $x=a$ the rank is μ.

The proof depends upon the following statements, (A), (B), and (C).

(A). The rank of Q will be $\mu-r$ if the linear equations

$$(11.2) \qquad Q_{u_\alpha} = 0 \qquad (\alpha = 1,2,\cdots,\mu)$$

have just r linearly independent solutions, and conversely.

(B). The equations (11.2) are necessary and sufficient conditions on a set (u) for such a set to determine a broken secondary extremal E' without corners at the points (9.1).

To prove (B) suppose (11.2) holds for some given set (u). Suppose u_α in (11.2) is the coördinate y_h of the point P_k in the set (9.1). We differentiate the integral in (10.3) with respect to u_α, integrate by parts, and obtain the following result, holding for the given set (u):

$$(11.3) \qquad Q_{u_\alpha} = \lim_{x=x_k^-} \Omega_{\eta_h'} - \lim_{x=x_k^+} \Omega_{\eta_h'} = 0.$$

In the limiting process the arguments in $\Omega_{\eta_h'}$ are those on E'. If now we hold k fast and let $h=1,2,\cdots,n$, (11.3) gives us n equations corresponding to the point $x=x_k$. It follows from (5.6) that these n equations are compatible only if E' has no corner at $x=x_k$.

Conversely equations (11.3) clearly hold for a given set (u) if the secondary extremal determined by (u) has no corners.

Thus the statement (B) is proved.

(C). A set of broken secondary extremals E' are linearly dependent or not according as the corresponding sets (u) are linearly dependent or not.

For if the coördinates (η) on q such curves E' satisfy a linear relation identically in x, then upon setting x successively equal to

$$(11.4) \qquad x_1, \cdots, x_p$$

we see that the corresponding sets (u) satisfy the same linear relation.

Conversely if a linear combination of q of the sets (u) gives the null set (0), and if these q sets determine q broken secondary extremals E' then the same linear combination of the coördinates (η) on the q curves E' will be zero at each point at which x has the values

$$a, x_1, \cdots, x_p, b, \tag{11.5}$$

and hence be identically zero, since no point of (11.5) is conjugate to its predecessor.

Thus (C) is proved.

It follows from Lemma 1 of § 8 that $x=b$ is a conjugate point of $x=a$ of the rth order, when and only when there are just r linearly independent secondary extremals which pass through the points $x=a$ and $x=b$ on the x axis. The theorem now follows from statements (C), (B), and (A).

12. **The type number of the fundamental form.** A real, symmetric, non-singular quadratic form,

$$Q = a_{\alpha\beta}u_\alpha u_\beta \qquad (\alpha,\beta = 1,2, \cdots, \mu), \tag{12.1}$$

can be transformed by a suitable real, linear, non-singular transformation into a form

$$-z_1^2 - \cdots - z_q^2 + z_{q+1}^2 + \cdots + z_\mu^2.$$

The integer q is called the *type number* of the given form.

LEMMA 1. *If the form Q is negative at each point of an r-plane π through the origin, excepting the origin, then the type number of Q is at least r.*

By a suitable non-singular linear transformation into variables (w) we first take the r-plane π into the r-plane

$$w_{r+1} = w_{r+2} = \cdots = w_\mu = 0. \tag{12.2}$$

After this transformation suppose Q takes the form

$$b_{\alpha\beta}w_\alpha w_\beta \qquad (\alpha,\beta = 1, \cdots, \mu). \tag{12.3}$$

On (12.2) Q will become a negative definite form

$$b_{\alpha\beta}w_\alpha w_\beta \qquad (\alpha,\beta = 1, \cdots, r). \tag{12.4}$$

Let A_m be that principal minor of the determinant of the form (12.3) which is obtained by striking out all but the first m rows and columns. Let $A_0=1$.

Note that A_r cannot be zero since the form (12.4) is definite. According to the well known theory of quadratic forms the variables $w_1, \cdots, w_r$ can be reordered so that no two consecutive members of the sequence

$$A_0, A_1, \cdots, A_r \tag{12.5}$$

are zero. So ordered, the number of changes of sign in (12.5) is r. The remaining variables of the set (w) can now be so reordered among themselves that no two consecutive numbers of the sequence

$$A_0, A_1, \cdots, A_\mu \tag{12.6}$$

are zero. So ordered, the number of changes of sign in (12.6) gives the type number of Q. Thus this type number is at least r and the lemma is proved.

LEMMA 2. *Suppose the coefficients of the form Q are continuous functions of a parameter t. If for $t=0$ the matrix $\mathbf{a}$ is of rank $\mu-r$, but is of rank μ for other values of t, then as t passes through $t=0$ the type number of Q can change in absolute value by at most r.*

According to the theory of symmetric quadratic forms the type number of Q will equal the number of λ roots of the characteristic matrix of $\mathbf{a}$ which are negative. These λ roots are continuous functions of the elements $a_{\alpha\beta}$ and hence of t. If $\mathbf{a}$ is of rank $\mu-r$ at $t=0$, just r of these roots will be zero. Hence r at most of these roots will change sign. The lemma follows directly.

LEMMA 3. *If $x=b$ is not conjugate to $x=a$, the type number of the form,*

$$Q = \sum J_{ij}u_iu_j \qquad (i,j = 1, \cdots, \mu) \tag{12.7}$$

is at most equal to the sum of the orders of the conjugate points of $x=a$ preceding $x=b$.

Consider the set of x coördinates

$$a, x_1, \cdots, x_p, b \tag{12.8}$$

of §9. To prove the lemma we shall hold a fast but vary the remaining coördinates in (12.8), using, however, only sets (12.8) which are admissible in the sense of §9.

Suppose the first conjugate point of $x=a$ is $x=a_1$. We start our variation of the coördinates (12.8) with a choice of $b<a_1$. For this choice of b and under the condition (3.3) the x axis from $x=a$ to $x=b$ furnishes a proper minimum* for the integral J, so that the type number q will be zero.

Suppose that a_1 is a conjugate point of a of the rth order. When $b=a_1$ the rank of the matrix of Q, according to Theorem 1, will be $\mu-r$. According to Lemma 2 of this section an increase of b through a_1 will be accompanied by an increase of the type number of Q at most r.

It is clear that a set of coördinates (12.8) for which $b<a_1$ can be varied into any other admissible set with the same a and the same p. One way of

* See Bliss, loc. cit., second paper.

doing this is first to increase x_1 to its final position keeping the following coördinates in (12.8) so near x_1 as to be admissible, then while holding a and x_1 fast, to increase x_2 to its final position, keeping the following coördinates in (12.8) so near x_2 as to be admissible, and so on.

As b thereby increases through each point conjugate to a prior to b's final position, there will be an accompanying increase of the type number of Q at most equal to the order of the conjugate point thereby passed. The lemma follows at once.

13. **The theorem on the type number of Q.** By an *auxiliary curve* g will be understood a finite succession of secondary extremals joining $(a, 0)$ to $(b, 0)$, whose points of junction are all on the x axis, and not all of which are segments of the x axis. An auxiliary curve g will intersect the successive hyperplanes

$$x = x_1, \cdots, x = x_p$$

in successive points P_i which we may suppose given in (9.1). But the points (9.1) determine the set (u) of §9. A set (u) thereby determined by an auxiliary curve g will be termed an *auxiliary set* (u).

No auxiliary set (u) is null.

This follows from the fact that each point of the set (11.5) precedes its predecessor's conjugate point. Further a set of auxiliary curves g will be linearly dependent if and only if the corresponding sets (u) are linearly dependent.

The following theorem is fundamental.

THEOREM 2. *In case $x=b$ is not conjugate to $x=a$ the type number q of the form*

$$Q = \sum J_{ij}u_iu_j \qquad (i,j = 1, \cdots, \mu) \tag{13.1}$$

will equal the sum of the orders of the conjugate points of a between $x=a$ and $x=b$.

Suppose the conjugate points of $x=a$ prior to $x=b$ have x coördinates

$$a_1 < a_2 < \cdots < a_s, \tag{13.2}$$

and their respective orders are

$$r_1, r_2, \cdots, r_s. \tag{13.3}$$

According to Lemma 3, §12, the type number q is such that

$$q \leqq r_1 + r_2 + \cdots + r_s. \tag{13.4}$$

We shall use Lemma 1 of §12 to prove that (13.4) must be an equality.

Corresponding to the ith conjugate point a_i, Lemma 1, §8, affirms the existence of r_i linearly independent secondary extremals,

$$h_j^{(i)} \qquad (j = 1, \cdots, r_i; \quad i = 1,2, \cdots, s), \tag{13.5}$$

which pass from $(a, 0)$ to $(a_i, 0)$. From the curves (13.5) we can now form r_i auxiliary curves,

$$g_j^{(i)}, \tag{13.6}$$

which are identical with the curves (13.5) as x varies from a to a_i, and are identical with the x axis as x varies from a_i to b.

Let

$$(u)_j^{(i)} \qquad (j = 1, \cdots, r_i; \quad i = 1,2, \cdots, s) \tag{13.7}$$

be the auxiliary sets (u) determined respectively by the curves (13.6). Our theorem will follow once we have proved the two following statements.

(A). *The $r_1+r_2+\cdots+r_s$ sets (u) in* (13.7) *are linearly independent.*

(B). *Any linear combination $(u)\neq 0$ of the sets* (13.7) *will make the fundamental form Q negative.*

To prove (A) it will be sufficient to prove the auxiliary curves (13.6) are linearly independent.

If there were any linear relation between the curves (13.6) which in particular involved those for which $i=s$, this would entail a linear relation among the curves for which $i=s$ alone, since the remaining curves of the set (13.6) coincide with the x axis between $x=a_{s-1}$ and $x=a_s$, and the curves for which $i=s$ do not. Similarly there can be no linear relation between the curves (13.6) which involves the curves for which $i=s-1$, and so on down to the curves for which $i=1$.

Thus (A) is proved.

We come to the proof of (B). Let (u) be any linear combination of the sets (13.7) not $(u)=(0)$. Let g be the auxiliary curve obtained by taking the same linear combination of the η-coördinates of the curves of the set (13.6). Let E' be the broken secondary extremal determined by (u) as in §10.

Without loss of generality we can suppose that none of the points at which x equals $x_1, \cdots, x_p$ are conjugate to $x=a$. For if such were not the case a slight displacement of these points would move them from positions conjugate to $x=a$, and would not alter the type number of Q. With this understood we see that the corners of g cannot coincide with any of the corners of E', since the corners of g are at the conjugate points of $x=a$, and the corners of E' are not.

I say now that E' will make the integral in (10.3) negative.

For g will make that integral zero since g is composed of secondary extremals with corners on the x axis.* But there will be some component secondary extremal of E', say E'', without any corner, which joins the end points of a portion g' of g with a corner. Under the condition (5.3), E'' will give a smaller value to the integral than g' so that E' must make the integral negative. Statement (B) now follows from (10.3).

To return to the theorem we note that the set of linear combinations (u) of the sets (13.7) may be regarded as the set of points on an

$$r_1 + r_2 + \cdots + r_s$$

plane through the origin in a space of the points (u). On this plane Q is negative except for the origin. The theorem follows from Lemma 1 of §12.

We saw in §8 that if $x=b$ is not conjugate to $x=a$, then $x=a$ is not conjugate to $x=b$. By interchanging the rôles of b and a in the preceding proof we can prove the following:

The type number of the form Q of Theorem 2 is also equal to the sum of the orders of the conjugate points of $x=b$, between $x=a$ and $x=b$.

We have thus proved the following:

If the point $x=b$ on γ is not conjugate to $x=a$, then $x=a$ is not conjugate to $x=b$, and the sum of the orders of the conjugate points of $x=a$ between $x=a$ and $x=b$ equals the sum of the orders of the conjugate points of $x=b$ between $x=a$ and $x=b$.

II. The fundamental form for the case of one variable end point

14. The hypotheses and definition of the form. Let O be any point in the region R of §2. Except for O suppose that R is covered in a one-to-one manner by the field of extremals through O, and that each extremal through O passes out of R before a first conjugate point of O is reached. We assume that F is *positively regular* along each extremal through O. In R let Σ be any closed regular† $(m-1)$-manifold of class C'''. We assume that $F(x, r)$ is positive for each point (x) on Σ, and for the direction (r) of the field at that point, or in particular if O lies on Σ, for all directions r at O. This concludes our statement of hypotheses.

We shall next set up the fundamental quadratic form. Let

$$x_i = h_i(v_1, \cdots, v_n) \qquad (i = 1, 2, \cdots, m)$$

---

* See Bolza, *Variationsrechnung*, p. 62, first formula.

† That is, a manifold the neighborhood of each point of which can be admissibly represented. See footnote §6.

be an admissible representation of Σ in the neighborhood of a point P of Σ. At P suppose $(v)=(v^0)$. Let the value of the integral J taken along the field extremal from O to the point (v) on Σ be denoted by

$$I(v_1, \cdots, v_n) = I(v) \qquad (n = m-1). \tag{14.1}$$

The function $I(v)$ will be of class* C'' in (v) for (v) neighboring (v^0). Suppose $I(v)$ has a critical point when $(v)=(v^0)$. The conditions that $I(v)$ have a critical point are the conditions that Σ cut the extremal g from O to P transversally. They are that

$$I_{v_j} = F_{r_i}(x,\dot{x})\frac{\partial h_i}{\partial v_j} = 0 \quad (j = 1,2, \cdots, m-1\,;\, i = 1,2, \cdots, m), \tag{14.2}$$

where (x) gives the coördinates of P, and $(\dot{x})$ is the direction of the field at P.

If then Σ cuts the extremal g transversally the terms of the first order in $I(v)$ at (v^0) all vanish. We next turn to the terms of the second order in $I(v)$, and in particular are concerned with the rank and type of the fundamental quadratic form,

$$Q = {}_0I_{ij}(v_i - v_i^0)(v_j - v_j^0) \qquad (i,j = 1,2, \cdots, n), \tag{14.3}$$

where the subscript zero indicates that the second partial derivatives I_{ij} are evaluated at $(v)=(v^0)$. This leads us to a study of focal points.

15. Focal† points and their orders. It is known that under our hypotheses‡ there exists a family of extremals neighboring g, passing through the points Q on Σ neighboring P in the senses of the extremals from O to Q, and cut transversally by Σ. They may be represented in the form

$$x_i = w_i(s,v) \qquad (i = 1,2, \cdots, m) \tag{15.1}$$

where s is the arc length taken as positive when measured along these extremals from Σ in their positive senses, and where the functions w_i and their first partial derivatives with respect to s are of class C'' for (v) neighboring (v^0), and for such values of s as give points neighboring g. The jacobian

$$\Delta(s) = \frac{D(w_1, \cdots, w_m)}{D(s,v_1, \cdots, v_n)}, \quad n = m-1, \quad (v) = (v^0) \tag{15.2}$$

does not vanish at $s=0$.

* This follows readily upon computing the first partial derivatives of $I(v)$ as in (14.2).

† An interesting paper on focal points has been written by M. B. White, *The dependence of focal points upon curvature for problems of the calculus of variations in space*, these Transactions, vol. 13 (1912), pp. 175–198.

‡ See Mason and Bliss, loc. cit., p. 448.

The points on g at which $\Delta(s)=0$ are called the focal points of P on Σ, and the order of vanishing of $\Delta(s)$ at each of these points the order of that point.

A necessary fact in what follows is that g cannot be tangent to Σ at P, as well as cut transversally by Σ at P. For the transversality conditions would then lead to the equation

$$F_{r_j}(x,\dot{x})\dot{x}_j = 0. \tag{15.3}$$

where (x) gives the coördinates of P, and $(\dot{x})$ the direction of g at P. But the left member of (15.3) equals F. Thus F would be zero contrary to our hypotheses on Σ.

16. **The reduction to non-parametric form.** As in §5, so here, the extremal g can be carried by an admissible transformation T into a segment γ of the x axis of a space of coördinates $(x, y_1, \cdots, y_n)$. Without loss of generality we can also suppose that T carries Σ into a manifold S that is orthogonal to γ at γ's final end point. Suppose g's end points O and P correspond to $x=a$ and $x=b$ on γ, respectively, with $a<b$. In the neighborhood of its intersection with γ, S can be represented in the form

$$x = \bar{x}(u_1, \cdots, u_n), \quad y_i = u_i \qquad (i = 1,2, \cdots, n), \tag{16.1}$$

where $\bar{x}(u)$ is of class C''' for (u) neighboring (0).

The transformation T establishes a one-to-one correspondence between the points determined by (v) on Σ, and the points determined by (u) on S, at least for (u) neighboring (0). This correspondence will take the form

$$v_i = h_i(u_1, \cdots, u_n)$$

where the functions $h_i(u)$ are of class C''' for (u) neighboring $(u)=(0)$.

By virtue of the transformation T the family of extremals neighboring γ, cut transversally by S, can be represented in terms of the variables s and (v) of (15.1) in the form

$$x = h(s,v), \quad y_i = k_i(s,v) \qquad (i = 1,2, \cdots, n). \tag{16.2}$$

As in the proof of the lemma in §6, so here, we can take x instead of s as a parameter along the extremals. We shall also replace the v_i's by their values $h_i(u)$ in term of the u_i's. The family (16.2) will then take the form

$$y_i = a^{(i)}(x,u) \qquad (i = 1,2, \cdots, n). \tag{16.3}$$

As in the proof of the lemma of §6, so here, a consideration of our successive transformations leads to the following results.

The focal points of Σ relative to the end point of g on Σ correspond under T to the zeros on γ of the jacobian

$$\frac{D(a^{(1)}, \cdots, a^{(n)})}{D(u_1, \cdots, u_n)} = \Delta_1(x), \quad (u) = (0), \tag{16.4}$$

and the orders of these focal points equal the orders of the corresponding zeros of $\Delta_1(x)$.

Under admissible space transformations focal points correspond to focal points, and their orders are preserved.

The columns of the jacobian (16.4) are solutions of the J.D.E. We shall find what conditions these columns satisfy at $x = b$.

The manifold $x = \bar{x}(u)$, $y_i = u_i$, cuts the family (16.3) transversally. We accordingly have the following identities in (u):

$$(f - p_h f_{p_h})\bar{x}_{u_i} + f_{p_i} \equiv 0 \qquad (i = 1, 2, \cdots, n), \tag{16.5}$$

in which we must set

$$x = \bar{x}(u), \; y_i = u_i, \; p_h = a_x^{(h)}[\bar{x}(u), u] \qquad (h = 1, \cdots, n).$$

Denote the partial derivatives of $\bar{x}(u)$ with respect to u_i and u_j by $\bar{x}_{ij}$. Now let $a_{ij}(x)$ be the general element of the jacobian (16.4). If we differentiate (16.5) with respect to u_j, and set $(u) = (0)$, we find that

$$f^0\bar{x}_{ij}(0) + B_{ik}(b)a_{kj}(b) + A_{ik}(b)a'_{kj}(b) = 0 \qquad (i, j, k = 1, \cdots, n), \tag{16.6}$$

where the superscript zero means that (x, y, p) is to be replaced by $(b, 0, 0)$. Further, differentiation of the identity

$$u_i \equiv a^{(i)}[\bar{x}(u), u]$$

with respect to u_i and u_j respectively yields the result

$$a_{ij}(b) = 1, \qquad i = j,$$
$$a_{ij}(b) = 0, \qquad i \neq j.$$

Consider now any linear combination of the columns of (16.4) of the form

$$\eta_k(x) = u_j a_{kj}(x) \qquad (k, j = 1, 2, \cdots, n)$$

where the u_j's are constant. Such a set takes on the values $(u_1, \cdots, u_n)$ when $x = b$. If we multiply (16.6) by u_j and sum with respect to j, we find that (η) satisfies the n relations,

$$f^0\bar{x}_{ij}(0)u_j + B_{ik}(b)\eta_k(b) + A_{ik}(b)\eta_k'(b) = 0 \qquad (i = 1, \cdots, n). \tag{16.7}$$

We conclude with the following lemma.

LEMMA. *Any linear combination of the columns of* (16.4) *which takes on the values* $(u_1, \cdots, u_n)$ *when* $x=b$, *will satisfy the n relations* (16.7) *at* $x=b$, *and as a solution of the J.D.E. be uniquely determined by these values* (u) *at* $x=b$, *and by the relations* (16.7).

17. **The rank of the fundamental form.** Consider again the function $I(v)$ of §14 which gives the value of the integral along the extremals from O to Σ. In terms of the parameters (u) which represent S in the neighborhood of the end point of γ, $I(v)$ becomes a function

$$J(u_1, \cdots, u_n) = J(u).$$

The function $J(u)$ gives the value of the transformed integral along the extremals from the point $x=a$ on γ to the point on S at which $(y)=(u)$. The form

$$Q = {}_0J_{ij}u_iu_j \qquad (i,j = 1,2, \cdots, n),$$

in which the subscript 0 indicates that the partial derivatives are evaluated at $(u)=(0)$, will have the same rank and type number as the form (14.3).

We shall now prove the following lemma.

LEMMA. *The fundamental form can be expressed as follows*:

$$(17.1) \qquad {}_0J_{ij}u_iu_j = f^0\bar{x}_{ij}(0)u_iu_j + \int_a^b \Omega(x,\eta,\eta')dx \qquad (i,j = 1, \cdots, n)$$

where (η) *stands for the coördinates* $[\eta_1(x), \cdots, \eta_n(x)]$ *on the secondary extremal joining* $x=a$ *on* γ *to the point* $(\eta)=(u)$ *on the n-plane* $x=b$.

Let the family of extremals which join the point $x=a$ on γ to the point at which $(y)=(u)$ on S, be represented in the form

$$y_i = m^{(i)}(x,u) \qquad (i = 1,2, \cdots, n),$$

where (u) lies in the neighborhood of $(u)=(0)$. As in §10, so here, we consider the function

$$(17.2) \qquad J(eu_1, \cdots, eu_n) = \int_a^{\bar{x}(eu)} f[x,m(x,eu),m_x(x,eu)]dx,$$

where e is a variable neighboring $e=0$.

If the two members of (17.2) be differentiated twice with respect to e, and then e be set equal to zero, (17.1) will result with

$$\eta_i(x) = u_j m_{u_j}{}^{(i)}(x,0).$$

That $\eta_i(b)=u_i$ follows upon differentiating the identity

$$u_i = m^{(i)}[\bar{x}(u), u]$$

with respect to u_j, and setting $(u) = (0)$.

The lemma is thereby proved.

The rank of Q is given by the following theorem.

THEOREM 3. *If $x = a$ on γ is a focal point of the rth order of S relative to the point $x = b$ on γ, then the rank of the matrix of the form Q is $n-r$. If $x = a$ is not a focal point of S the form is non-singular.*

As in the proof of Theorem 1, so here, we see that the rank of Q is $n-r$ if the n linear equations

$$Q_{u_i} = 0 \qquad (i = 1, 2, \cdots, n) \tag{17.3}$$

have just r linearly independent solutions (u) and conversely.

If we differentiate the right hand member of (17.1) with respect to u_i, and perform the usual integration by parts, we find that when (17.3) holds for a set (u), the functions (η) of (17.1) which take on at $x = b$ these values (u), satisfy (16.7). According to the lemma of §16 these functions (η) must be dependent on the columns of the jacobian of (16.4). But the functions (η) of (17.1) vanish when $x = a$.

We see that the number of linearly independent solutions (u) of the equations (17.3) equals the number of linearly independent solutions of the J.D.E. which are dependent on the columns of Δ_1, and which vanish at $x = a$. This number equals the order of vanishing of $\Delta_1(x)$ at $x = a$ and thus equals the order of the focal point $x = a$. Thus the theorem is proved.

18. **The type number of the form Q.** We shall now prove the following lemma.

LEMMA. *If $x = a$ on γ is not a focal point of S relative to the point $x = b$ on γ, the type number of the form Q is at most equal to the sum of the orders of the focal points of S on γ between $x = a$, and $x = b$.*

To prove this lemma we vary a from a value near $b, a < b$, to its given value. For $x = a$ nearer $x = b$ than any focal point of S, the form Q will be positively definite, since (5.3) holds along γ. As $x = a$ decreases through a focal point of order r there will be an accompanying increase of the type number of Q at most r. This follows from Lemma 2 §12 and Theorem 3. The present lemma follows directly.

The following theorem will now be proved.

THEOREM 4. *If $x = a$ on γ is not a focal point of S relative to $x = b$ on γ, the type number q of the form Q will equal the sum of the orders of the focal points of S between $x = a$ and $x = b$.*

The proof is similar to the proof of Theorem 2. The following differences should however be noted.

By an *auxiliary curve* $y_i = \eta_i(x)$ we shall here mean a curve which is made of two secondary extremals as follows. The first shall consist of the x axis from $x=a$ to a point $x=a_1$ between $x=a$ and $x=b$. The second shall pass from the point $x=a_1$ on the x axis to a point on S at which $(y)=(u)\neq(0)$, and shall be dependent on the columns of Δ_1 in (16.4).

If in (17.1) (η) be taken as the coördinates along an auxiliary curve, and if u_i be taken as $\eta_i(b)$, the right hand member of (17.1) is zero. This is readily seen if the integral in (17.1) be written in the form

$$\frac{1}{2}\int_a^b (\eta_i\Omega_{\eta_{i'}} + \eta_{i'}\Omega_{\eta_{i'}})dx,$$

and the second sum be integrated by parts. The resulting integral vanishes. The right member of (17.1) then becomes

$$f^0x_{ij}(0)u_iu_j + \tfrac{1}{2}u_i\Omega_{\eta_{i'}}[b,\eta(b),\eta'(b)]. \tag{18.1}$$

This expression may also be obtained by multiplying the left member of (16.7) by u_i, and summing with respect to i, and is accordingly zero.

It follows that any secondary extremal E that joins the end points of an auxiliary curve will make the right member of (17.1) negative, since E will give the integral in (17.1) a value that is smaller than that given by the auxiliary curve.

If $x=a_1$ is a focal point of order r there will be r linearly independent auxiliary curves with corners at $x=a_1$ on γ. We now continue as in the proof of Theorem 2, and complete the proof of the present theorem.

III. Theorems in the large on extremals through a fixed point

19. **The function $I(v)$.** We return to the manifold Σ of §14 and to the field of extremals through O. We assume *that O is not a focal point of* Σ. The locus of focal points at a finite distance from Σ will form in general a finite number of manifolds of dimensionality at most $m-1$. To take O therefore as a point which is not a focal point is to take the general point O.

The integral J taken along the extremals of the field from O to a point Q on Σ will be a function I of the point Q on Σ. The critical points of I are to be determined by expressing I in the neighborhood of each point P on Σ in terms of admissible local parameters (v). Since O is not a focal point on Σ these critical points will be non-degenerate, that is, points at which the fundamental quadratic form (14.3) will be of rank n. It follows that these critical points are isolated on Σ and therefore finite in number. Their type

numbers will be given by Theorem 4 of §18, and will be called the type numbers of the corresponding extremals of the field.

We shall now prove the following theorem.

THEOREM 5. *Corresponding to a continuous variation of a point O' in the neighborhood of O, the end points of the extremals from O' to Σ which are cut transversally by Σ will vary continuously on Σ while the corresponding type numbers will be unchanged.*

In §14 we imposed conditions on the field of extremals through O. If O' be restricted to a sufficiently small neighborhood of O, the field of extremals through O' will satisfy these same conditions. Let $(x) = (a)$ be the coördinates of O', and $(x) = (a^0)$ the coördinates of O. Let (v) be a set of parameters in an admissible representation of Σ in the neighborhood of the point P. The integral taken along the extremal from O' to a point (v) on Σ will be a function $I(v, a)$ of (v) and (a).

The conditions that $I(v, a)$ have a critical point when considered as a function of (v) alone are

$$I_{v_i}(v, a) = 0 \qquad (i = 1, 2, \cdots, n). \tag{19.1}$$

Suppose the extremal from O to P is cut transversally by Σ at P. Let $(v) = (v^0)$ correspond to P. The equations (19.1) will have an initial solution (v^0, a^0). The jacobian of the functions I_{v_j} with respect to the parameters (v) is

$$| I_{v_i v_j} | .$$

According to Theorem 3 of §17 this jacobian will not be zero when $(v, a) = (v^0, a^0)$, since O is not a focal point of Σ.

The equations (19.1) accordingly determine the parameters (v) as single-valued continuous functions of the point (a) at least for positions of (a) neighboring (a^0). Thus the end points of the extremals from O' to Σ which are cut transversally by Σ vary continuously on Σ as required.

That the type numbers of these extremals remain unchanged follows from the fact that the coefficients of the fundamental form (14.3) can now be regarded as continuous functions of (a).

20. **The fundamental theorem in the large.** We need the following lemma.

LEMMA. *At a point $(x) = (a)$ at which both $F(a, r)$ and $F_1(a, r)$ vanish for no direction (r), there will exist, corresponding to any regular $(m-1)$-dimensional manifold S through (a), at least one direction cut transversally by S at (a).*

Reference to the transversality conditions shows that a direction $(\dot{x})$ will be cut transversally by S at $(x)=(a)$, if it is cut transversally by the $(m-1)$-plane π tangent to S at $(x)=(a)$.

Let (λ) be any direction. The conditions of transversality show that there exists at least one $(m-1)$-plane ω through (a) which cuts (λ) transversally. Now let ω be continuously turned about (a) so as to coincide with π. That we can solve the transversality conditions for the directions (λ) in terms of the parameters which determine the variable position of ω follows upon consideration of a jacobian set up by Mason and Bliss.* Although this jacobian is not used by Mason and Bliss for exactly our purpose, still they show that it does not vanish if F and F_1 are not zero.

Thus as ω is continuously turned into the plane π the direction (λ) will continuously turn to a direction that is cut transversally by π. The lemma is thereby proved.

If O is not a focal point of Σ the critical points of the function I of §14 are non-degenerate. Let M_k be the number of critical points of type k. Let R_i be the ith connectivity of Σ. According to a theorem proved by the author the following relations hold:†

$$
\begin{aligned}
M_0 &\geqq 1 + (R_0 - 1), \\
M_0 - M_1 &\leqq 1 + (R_0 - 1) - (R_1 - 1), \\
M_0 - M_1 + M_2 &\geqq 1 + (R_0 - 1) - (R_1 - 1) + (R_2 - 1), \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
M_0 - M_1 + \cdots + (-1)^n M_n &= 1 + (R_0 - 1) - + \cdots + (-1)^n (R_n - 1),
\end{aligned}
\tag{20.1}
$$

where the inequality signs alternate until the final equality is reached.

By virtue of Theorem 4, §18, the preceding statement may be put into the following form.

THEOREM 6. *Suppose O is not a focal point of Σ nor on Σ. Of the extremal segments from O to Σ which are cut transversally by Σ at their ends P on Σ let M_k be the number upon which the sum of the orders of the focal points of P between O and P is k. Then between the numbers M_k and the connectivities R_i of Σ the relations* (20.1) *hold.*

If O is on Σ the above relations still hold if we count O as if it were an extremal of type zero.

---

* Loc. cit., p. 447.

† Marston Morse, *Relations between the critical points of a real function of n independent variables*, these Transactions, vol. 27 (1925), p. 392. The theorem proved in this reference admits an obvious generalization to closed manifolds such as Σ.

The theorem is already established except in the special case where O is at a point (a) on Σ. In this case it follows from the lemma of this section that there is at least one extremal E issuing from Σ at (a) which is cut transversally by Σ at (a). If we now take O in a position on E not (a) but sufficiently near (a), the extremal from O to (a) will be of type zero. According to the proof of Theorem 5, §19, the other extremals from O to Σ which are cut transversally by Σ will be changed at most slightly in position, but will be unchanged in type and number. Thus the theorem follows as stated when O is on Σ.

The following corollary furnishes a strong existence theorem.

COROLLARY. *If O is not a focal point of Σ, the number of extremals from O to Σ cut transversally by Σ on which the sum of the orders of the focal points is k is at least R_k-1. The total number of extremals is at least*

$$1 + (R_0 - 1) + \cdots + (R_n - 1).$$

Even when O is a focal point of Σ one can use a limiting process to get strong existence theorems. Considerations of this sort will be presented in a later paper.

21. **Normals from a point to a manifold.** A beautiful geometric application of these results obtains if we consider the integral which gives the arc length. The hypotheses of §14 are here automatically fulfilled for any point O. It remains to consider the focal points.

Let P be an arbitrary point on Σ. Let P be taken as the origin, and the $(m-1)$-plane tangent to Σ at P be taken as the $(m-1)$-plane $x_m=0$. After a suitable rotation of the remaining axes in the $(m-1)$-plane $x_m=0$, Σ may be represented as follows:

$$2x_m = b_i x_i^2 + H \qquad (i = 1,2, \cdots, m-1), \tag{21.1}$$

where b_i is a constant, and H is of the third order in $(x_1, \cdots, x_{m-1})$.

For such of the constants b_i as are not zero we set

$$R_i = \frac{1}{b_i},$$

and call the point P_i on the x_m axis at which $x_m=R_i$, the ith center of a principal normal curvature of Σ corresponding to P.

If a constant $b_i=0$, we say that the corresponding center P_i is at infinity. We define the centers P_i for any axes obtained from our specialized set by a rotation or translation, by imagining each normal to Σ rigidly fixed to Σ, and each center P_i rigidly fixed to its normal.

To determine the focal points on the normal to Σ at P we take the parameters (v) of §14 as the set $(x_1, \cdots, x_{m-1})$, and measure s along the normals in the sense of increasing x_m. We make use of the representation

(21.1) to obtain the family of normals to Σ at points neighboring P. The jacobian $\Delta(s)$ of §15 evaluated for $(v)=(v^0)$ is here seen to have the form

$$\Delta(s) = (1 - b_1 s)(1 - b_2 s) \cdots (1 - b_{m-1} s).$$

Thus the sum of the orders of the focal points on a normal to Σ, between the foot P of that normal on Σ and a point O on the normal, equals the number of centers of principal normal curvature between P and O.

It is understood that for the purpose of counting the number of centers P_i between P and O, two centers P_i and P_j are to be counted separately if $i \neq j$, even if P_i and P_j are identical in position. In case O is on Σ it is also understood that O is to be counted as a single normal on which there are, of course, no centers P_i between O and P. Finally two normals from O to Σ which have the same direction are to be counted as distinct if they have different feet on Σ. With this understood Theorem 6 gives us the following:

THEOREM 7. *Let O be a point that is not a center of principal normal curvature of Σ. Of the normals from O to Σ let M_i be the number of those on which there are i centers of principal normal curvature of Σ between O and the feet of the normals on Σ. Then between these numbers M_i and the connectivities R_i of Σ the relations* (20.1) *hold.*

COROLLARY. *If O is not a center of principal normal curvature of Σ, the number of distinct normals from O to Σ on which there are k centers of principal normal curvature is at least $R_k - 1$. The total number of distinct normals is at least*

$$1 + (R_0 - 1) + \cdots + (R_n - 1).$$

22. **Examples.** Let Σ be a regular surface of genus p in three-space. Corresponding to a point O which is not a center of principal normal curvature of Σ we have

$$M_0 \geqq 1, \quad M_1 \geqq 2p, \quad M_2 \geqq 1.$$

Thus there will be at least $2p+2$ distinct normals from O to Σ. In addition we have the relation

$$M_0 - M_1 + M_2 = 2 - 2p.$$

Or again let Σ be a manifold in 4-space homeomorphic to a manifold obtained by identifying the opposite faces of an ordinary cube. If O is not a center of principal normal curvature of Σ we have

$$M_0 \geqq 1, \quad M_1 \geqq 3, \quad M_2 \geqq 3, \quad M_3 \geqq 1.$$

There are thus at least eight normals from O to Σ.

HARVARD UNIVERSITY,
CAMBRIDGE, MASS.

CERTAIN INVARIANT SEQUENCES OF POLYNOMIALS*

BY

E. T. BELL

We discuss certain sequences of uniform functions of one variable such that the derivative of each term of a given sequence is equal to the preceding term, and each term is changed by a linear transformation on the variable into a multiple of itself, the multiplier being a function of the rank of the term alone, and the linear transformation and the multiplier function being the same for all terms. It is an interesting problem to determine all such sequences of a certain type, described presently, and to assign the corresponding transformations and multipliers. This is done in the following sections; there is an infinity of solutions. Although we consider only functions of one variable, the method is general and applicable to functions of any number of variables.

The terms are necessarily polynomials. By linear transformations on the rank and the variable, any number of distinct sequences having the stated properties for any given linear transformations on their variables can be replaced by new sequences, all of which are transformed alike by the same linear transformation on the variable, and hence are instances of a single sequence of the original kind. In this respect the theory of any number of distinct sequences of the kind described can be unified.

The determination of *all* sequences of the general kind just mentioned, presents no difficulty. The solution, if no restriction be imposed on the numbers defined by the sequences, contains an infinity of arbitrary constants. From our point of view, which is to extend in as simple a manner as possible the existing instances of such sequences, the perfectly general solution is of but slight interest; it is difficult to see how it could lead to interesting extensions of the known cases. This remark is the origin of §4, Theorem 3, where the alternative condition $\psi_0\psi_2-\psi_1^2=0$ is rejected, and its contradictory leads to a definite infinity of solutions, each of which contains only a finite number of arbitrary constants. We have endeavored to construct the theory so that its interest shall be arithmetical rather than algebraic.

There are four classic instances of such sequences. These will be derived in §8 as immediate special cases, to provide checks on the general theorems. The main points of the paper are the definitions of index and characteristic in §1, and the Theorems 1–11.

* Presented to the Society, San Francisco Section, October 27, 1928; received by the editors in September, 1928.

Both in the classic instances and in their extensions, the sequences appear as particular solutions of what is here called the functional equation of invariant generators. This functional equation alone is not sufficient to define a particular sequence. The manner in which particular solutions are completely specified is explained in §§8, 9. For example, the Bernoulli numbers being the coefficients in the successive polynomials of one classic sequence, we use the even suffix notation $B_0, B_2, B_4, B_6, \cdots$, in which, if no further condition be imposed, $B_1, B_3, B_5, \cdots$ are entirely arbitrary. To eliminate this undesirable infinity of arbitrary constants, we adjoin to the functional equation another for the numbers B_{2n} $(n=0, 1, \cdots)$, in a perfectly definite manner, which completely defines the sequence, and so in all cases.

1. **Invariant sequences.** Let n be an integer $\geqq 0$, and x a real or complex variable. A statement involving n shall signify the totality of statements obtained from the given one by taking $n=0, 1, 2, \cdots$, successively, so that it will be unnecessary in formulas and elsewhere to indicate the range of n.

The notation $(a_1, \cdots, a_r)=(b_1, \cdots, b_r)$ means $a_i=b_i (i=1, \cdots, r)$; while $(a_1, \cdots, a_r)\neq(b_1, \cdots, b_r)$ means that at least one of $a_i=b_i$ $(i=1, \cdots, r)$ is false.

Let $f_n(x)$ be a single-valued function of x defined for all x as above, and, when necessary, impose the convention that $f_{-1}(x)$ is defined and finite for all values of x considered. If $f_n(x)$ is a polynomial in x, its degree in x is by definition n. If y is a function of x, the derivative of $f_n(y)$ with respect to x will be denoted by $f_n'(y)$. It is necessary to assume that $f_n'(y)$ exists only in what immediately follows; thereafter $f_n'(y)$ automatically exists.

It is well known and indeed obvious that the general solution of

$$f_n'(x) = nf_{n-1}(x) \tag{1}$$

is the polynomial

$$f_n(x) = (x+\alpha)^n \equiv \sum_{s=0}^{n} \binom{n}{s} \alpha_{n-s} x^s,$$

where, as indicated, $(x+\alpha)^n$ is the *symbolic nth power* of $x+\alpha$, and α is the *umbra* of the sequence $\alpha_0, \alpha_1, \cdots, \alpha_n, \cdots$ of arbitrary constants. That is, the general solution of (1) is the *Appell polynomial* of *rank* n with α as *base*. Note that the coefficient of x^n in $f_n(x)$ is α_0. We may refer to α as the *base* of the sequence $f_0(x), f_1(x), \cdots, f_n(x), \cdots$.

Let h be independent of both n and x, and let $\gamma(n)$ be a function of n alone such that $\gamma(0)\neq 0, \infty$; the excluded values lead only to trivialities.

Then if $g_n(x)$ is any solution of (1), the necessary and sufficient condition that $\gamma(n)\, g_n(x)$ shall be a solution of

$$\psi_n'(x) = \psi_{n-1}(x), \tag{2}$$

is $n\gamma(n) = \gamma(n-1)$; hence $\gamma(n) = \gamma(0)/n!$, and we have the following:

If k is an arbitrary constant which takes neither of the trivial values 0, ∞, and if $f_n(x) \equiv g_n(x)$ is the general solution of (1), *the general solution of* (2) *is*

$$\psi_n(x) \equiv kg_n(x)/n!\,; \tag{3}$$

conversely, the general solution of (1) *is* $f_n(x) \equiv n!\,\psi_n(x)/k$.

We shall call $\psi_n(x)$ the *canonical* polynomial of *degree* n with *base* α, and, when necessary, say that $\psi_n(x)$ *corresponds to* the Appell polynomial $g_n(x)$ from which it is constructed, and refer to α as the *base of the sequence* $\psi_0(x), \psi_1(x), \cdots, \psi_n(x), \cdots$.

Let $\tau(n)$, called the *multiplier*, be a function of n alone which is finite for all integers $n \geqq 0$, and not identically zero. Then, if there exist constants a, b, other than the trivial pair $(a, b) = (1, 0)$, such that

$$f_n(ax + b) = \tau(n) f_n(x), \tag{4}$$

we shall call

$$f_0(x),\ f_1(x), \cdots, f_n(x), \cdots$$

an *invariant sequence with respect to the transformation* $[x, ax+b]$, or briefly, an *invariant sequence*. When there can be no confusion between the term $f_n(x)$ and the sequence of which this is the nth term, we shall refer to the sequence as $f_n(x)$. Note particularly that in this definition the terms are not restricted to be polynomials.

If $f_n(x)$ for a particular $(a, b, \tau(n))$ is the general solution of (4), then $kf_n(x)/n!$, where k is an arbitrary constant, is also a solution. Hence the simultaneous solutions, if any, of (2), (4) are canonical polynomials. Let $f_n(x)$ for a particular $(a, b, \tau(n))$ be the general solution of

$$f_n'(x) = f_{n-1}(x), \qquad f_n(ax + b) = \tau(n) f_n(x). \tag{5}$$

Then we define

$$f_0(x),\ f_1(x), \cdots, f_n(x), \cdots$$

to be an *invariant sequence of polynomials* with the *characteristic* $(a, b, \tau(n))$, or simply an *invariant polynomial sequence*, when the characteristic is understood or otherwise indicated.

Our problem is to determine all invariant polynomial sequences. This will be accomplished when we find the multiplier $\tau(n)$, the constants a, b in

the transformation $[x, ax+b]$, and sufficiently define the bases of the Appell polynomials corresponding to the canonical polynomials concerned for a given characteristic. We shall first partially dispose of the multiplier. This depends upon the index, next defined, which is the taproot of the whole theory. When b is an integer, the sequences are of particular interest (§6, Theorem 10).

Let ϕ be the umbra of the sequence of absolute constants $\phi_0, \phi_1, \cdots, \phi_n, \cdots$. The least integer $s \geqq 0$ such that $\phi_s \neq 0$, will be called the *index* of ϕ.

If ϕ is the base of the sequence $f_n(x)$, and s is the index of ϕ, we replace n by $n+s$ in (5), differentiate the result n times successively, and get

$$a^n f_s(ax+b) = \tau(n+s) f_s(x).$$

Since $f_s(x) \equiv (x+\phi)^s = \phi_s = f_s(ax+b)$, we have $\tau(n+s) = a^n$. Again, since $f_j(x) = 0$ $(j = 0, 1, \cdots, s-1)$, it is immaterial what finite values be assigned to $\tau(j)$ $(j = 0, 1, \cdots, s-1)$. In particular we may take $\tau(j) = a^{j-s}$ $(j = 0, 1, \cdots, s-1)$. Hence

$$\tau(n) \equiv a^{n-s} \tag{6}$$

is the value of the multiplier for the invariant polynomial sequence defined in (5), *in which* $f_n(x)$ *has base* ϕ, *and* s *is the index of* ϕ.

The classic instances of (5) are given by the polynomials whose bases are B, G, E, L, these being the umbrae of the sequences of the numbers of Bernoulli, Genocchi, Euler and Lucas. The customary manner of proving that these polynomials are indeed instances is somewhat fortuitous and effectively conceals the root of the matter, which is the index of the numerical sequence concerned in each case. Incidentally our general theorems give much more than the classic results for these instances. It will be interesting to observe the fundamental part played by the index in the general theory and in its applications to the classic instances.

2. **Equivalent sequences.** This section refers to the functional equation (4) in §1, so that $f_n(x)$ is not restricted to be a polynomial, and (6) does not necessarily hold.

The invariant sequences $f_n(x)$, $g_n(x)$ are defined to be *identical* if and only if $f_n(x) = g_n(x)$; otherwise they are *distinct*. Distinct sequences invariant with respect to the same given transformation $[x, ax+b]$ will be called *equivalent*. We now assign necessary and sufficient conditions for invariant sequences to be equivalent, and give the requisite formulas.

Let $a \neq 0$, $\alpha \neq 0$, b, β be constants other than $(a, b) = (\alpha, \beta) = (1, 0)$; the excluded values give only trivialities. Let r, s be constant integers $\geqq 0$, and

$f_n(x)$, $h_n(x)$ functions of x defined as in §1, beginning; also let $\tau(n)$ be as in (4), and $\sigma(n)$ a similarly defined function such that

$$f_n(ax+b) = \tau(n)f_n(x), \quad h_n(ax+b) = \sigma(n)h_n(x).$$

Let $c \neq 0$, $\gamma \neq 0$, d, δ be constants. If $\tau(n+r) \neq 0$, $\sigma(n+s) \neq 0$, we define $g_n(x)$, $k_n(x)$ by

$$g_n(x) \equiv f_{n+r}(cx+d), \qquad k_n(x) \equiv h_{n+s}(\gamma x+\delta).$$

Then, as may be easily verified, the functional equations (4) for g, k are

$$g_n\left(\frac{x}{a} + \frac{d-b-ad}{ac}\right) = \frac{1}{\tau(n+r)}g_n(x),$$

$$k_n\left(\frac{x}{\alpha} + \frac{\delta-\beta-\alpha\delta}{\alpha\gamma}\right) = \frac{1}{\sigma(n+s)}k_n(x).$$

We require the conditions upon the functions τ, σ and the several constants which shall yield two or more of the sequences

$$f_n(x), \ g_n(x), \ h_n(x), \ k_n(x)$$

as solutions of a single functional equation of the type

$$\xi_n(\lambda x+\mu) = \rho(n+j)\xi_n(x),$$

in which j is a constant integer $\geqq 0$, $\rho(n+j)$ is defined, finite and not identically zero for all integers $n \geqq 0$; λ, μ are constants other than $(\lambda, \mu) = (1, 0)$, and $\xi_n(x)$ is single-valued and finite for all x considered.

Comparing the above functional equations for f, g, h, k, we find that there are precisely two distinct non-trivial solutions. Writing $(\alpha, \beta, \delta) \equiv (a, p, t)$ in the solutions thus found, we get the following:

THEOREM 1. *If $ap \neq 0$, and $\tau(n+r) \neq 0$, the equation*

$$\xi_n\left(\frac{x}{a}+p\right) = \frac{1}{\tau(n+r)}\xi_n(x)$$

has the solution

$$\xi_n(x) = g_n(x) \equiv f_{n+r}\left(\frac{[d(1-a)-b]x}{ap} + d\right),$$

where d is an arbitrary constant, r an arbitrary constant integer $\geqq 0$, and $f_n(x)$ is any solution of

$$f_n(ax+b) = \tau(n)f_n(x).$$

THEOREM 2. *If* $ac \neq 0$ *and* $d(1-a) \neq b$, *and if and only if* $\tau(n+r) = \sigma(n+s) \neq 0$, *where* r, s *are integers* $\geqq 0$, *the equation*

$$\xi_n\left(\frac{x}{a} + \frac{d(1-a)-b}{ac}\right) = \frac{1}{\tau(n+r)}\xi_n(x)$$

has the solutions

$$\xi_n(x) = g_n(x) \equiv f_{n+r}(cx+d),$$

$$\xi_n(x) = k_n(x) \equiv h_{n+s}\left(\frac{c[t(a-1)+p]x}{d(a-1)+b} + t\right),$$

where d, t *are arbitrary constants and* $f_n(x)$, $h_n(x)$ *are any solutions of* $f_n(ax+b) = \tau(n)f_n(x)$, $h_n(ax+p) = \sigma(n)h_n(x)$.

Thus if in Theorem 1 a solution of each of the equations for ξ, f be known, g is a second solution of the ξ equation; if in Theorem 2 a solution for each of the f, h equations be known, the ξ equation has the two solutions g, k. The solutions can be easily verified. In §5 we find the equivalents of these theorems for invariant polynomial sequences and show how they are to be applied.

3. **Generators.** If $\phi_0, \phi_1, \cdots, \phi_n, \cdots$ is a sequence of numbers, real or complex, and z is a parameter, we shall call

$$e^{\phi z} \equiv \sum_0^\infty \phi_n(2^n/n!)$$

the *generator* of the sequence whose umbra is ϕ. The generator of the sequence of Appell polynomials in x with base ϕ is $e^{xz}e^{\phi z}$, or $e^{(x+\phi)z}$. Generators $e^{\phi z}$, $e^{\psi z}$ are defined to be *equal*, $e^{\phi z} = e^{\psi z}$, when and only when $\phi_n = \psi_n$. The symbolic or umbral calculus of such generators being well known, we may dispense with further details, except to remark that this calculus has been founded postulationally on an algebraic basis which renders all discussion of convergence in the use of generators for deriving relations between elements of sequences irrelevant.

Let $f_n(x)$ be the sequence of Appell polynomials with ϕ as base. Then

$$e^{(x+\phi)z} = e^{f(x)z}.$$

Let $g_n(x)$ be the sequence of canonical polynomials corresponding to $f_n(x)$. If there exist constants a, b, c, k, other than the trivial sets $(a, b, c, k) = (1, 0, 1, 1), (0, 0, 0, 0)$, such that

$$g_n(ax+b) = kc^n g_n(x),$$

we shall call $e^{\phi z}$ an *invariant generator*. If s is the index of ϕ, by §1(5), (6) we have

$$kc^n = \tau(n) = a^{n-s};$$

but it is more convenient in §4 to use kc^n instead of either of its equivalents $\tau(n)$, a^{n-s}. From the definitions in §1 it follows at once that the problem of determining all invariant polynomial sequences is identical with that of finding all invariant generators.

4. **Invariant generators.** Let u, v be independent variables, and $T(u, v)$ a function of u, v such that $T(z, e^z)$ is a generator as defined in §3. Let the index of ϕ be s; define ψ by

$$(n+s)!\psi_n \equiv n!\phi_{n+s},$$

and let the generator of ψ be $T(z, e^z)$. Then the index of ψ is zero, and

$$T(z,e^z) = e^{\psi z}, \qquad z^s T(z,e^z) = e^{\phi z}.$$

If now there exist constants (a, b, c, k) different from $(1, 0, 1, 1)$, $(0, 0, 0, 0)$ such that

$$e^{(ax+b)z}T(z,e^z) = kc^s e^{cxz}T(cz,e^z),$$

then and only then is $e^{\phi z}$ an invariant generator, as is evident on comparing the generators of $f_n(ax+b)$, $kc^n f_n(x)$, the notation being as in §3. Multiply throughout by e^{-cxz}. Then the new left member must be independent of x, since the new right is. Hence $a=c$, and we have

$$e^{bz}T(z,e^z) = ka^s T(az,e^{az}),$$

as the necessary and sufficient condition upon $T(u, v)$ in order that $T(z, e^z)$ shall be an invariant generator. From the last,

$$(\psi + b)^n = ka^{n+s}\psi^n.$$

Conversely, this implies the preceding equality, and hence it also is necessary and sufficient.

Reject the trivial cases $ka=0$. Take $n=0, 1, 2$ in the last. Then we get

$$ka^s = 1, \qquad b = (a-1)\psi_1/\psi_0, \qquad (a^2-1)(\psi_0\psi_2 - \psi_1^2) = 0.$$

Excluding $a=1$, which yields merely the identical transformation $[x, x]$, and noting, as is easily seen, that $\psi_0\psi_2 - \psi_1^2 = 0$ leads only to trivialities, we get the unique solution

$$(a,b,k) = (-1, -2\phi_{s+1}/[(s+1)\phi_s], (-1)^s).$$

THEOREM 3. *There exists precisely one non-trivial characteristic* $(a, b, \tau(n))$ *such that the sequence of canonical polynomials* $f_n(x)$ *with the base* ϕ *has the property*

$$f_n(ax + b) = \tau(n) f_n(x) ;$$

if s *is the index of* ϕ *the characteristic is*

$$(a, b, \tau(n)) = (-1, -2\phi_{s+1}/[(s+1)\phi_s], (-1)^{n+s}),$$

and ϕ *is generated by* $R(z, e^z)$, *where* $R(u,v)$ *is any solution of the functional equation*

$$v^b R(u,v) = (-1)^s R(-u, v^{-1}),$$

in which u, v *are independent variables and* b *is as above.*

From this we have

COROLLARY 1. *When the index* s *of* ϕ *is given, the first* $s+2$ *terms of* ϕ *are necessary and sufficient to determine the transformation with respect to which the invariant polynomial sequence with base* ϕ *is invariant, and its characteristic.*

COROLLARY 2. *A particular invariant polynomial sequence and its characteristic are uniquely determined by the generator of the base of the polynomials; conversely, a particular generator determines precisely one invariant polynomial sequence and its unique characteristic.*

The distinction between the information furnished by these two corollaries may be emphasized: the functional equations §1 (5), for a given characteristic $(a, b, \tau(n))$, are determined by the first $s+2$ terms only of the base, and have an infinity of solutions; to select from this infinity a particular solution it is necessary to know the generator of the base, not merely the first $s+2$ terms of the sequence which it generates.

To obtain an *element* of the solution of the functional equation of invariant generators, we assume that $R(u, v)$ is a sum of terms of the form $v^{\lambda b+\mu} A(u)$, where λ, μ are constants and $A(u)$ is independent of v.

THEOREM 4. *The functional equation of invariant generators*

$$v^b R(u,v) = (-1)^s R(-u, v^{-1}),$$

in which u, v *are independent variables,* b *is an arbitrary constant and* s *is an integer* $\geqq 0$, *has the elementary solution, involving both* u *and* v, *or only* v,

$$R(u,v) \equiv v^{rb+t} F(u) + v^{-(r+1)b-t}(-1)^s F(-u),$$

in which r, t *are arbitrary constants and* $F(u)$ *is an arbitrary function of* u *alone, including the case* $F(u)$ *constant.*

By requiring given rational functions of $v, v^{rb}, v^{tb}, \cdots$ and solutions of the equation to be further solutions, we easily obtain the following:

THEOREM 5. *All rational functions of given solutions of the functional equation of invariant generators that contain both variables u, v and are again solutions, can be constructed by repetitions of the operations indicated in*

$$kR(u,v), \quad R_1(u,v) + R_2(u,v),$$

$$(1 + (-1)^s)v^{sb/2}R_1(u,v)[R_2(u,v)]^s \qquad (\epsilon = 1, -1),$$

where b, s are as in Theorem 4, *k is an arbitrary constant, and* $R(u,v)$, $R_1(u,v)$, $R_2(u,v)$ *are given solutions.*

In applying this and the next, the usual precautions regarding vanishing functions as divisors are to be observed. The separable solutions are of some interest, as the four classic instances of invariant sequences mentioned in §1 have generators of this type.

THEOREM 6. *All solutions of the functional equation of invariant generators of the type* $A(u)\,B(v)$, *where* $A(u), B(v)$ *are functions of u alone, v alone, are given by*

$$A(u) = F(u) + (-1)^{\eta}F(-u), \qquad B(v) = G(v) \qquad (\eta = 0,1),$$

where $F(u)$ *is an arbitrary function of u, and* $G(v)$ *is any solution of*

$$G(v^{-1}) = (-1)^{(\eta+1)s}v^{b}G(v),$$

where the same value of η *is to be used in both of* $A(u)$, $B(v)$. *The G equation has the elementary solution*

$$G(v) \equiv v^{rb+t} + (-1)^{(\eta+1)s}v^{-(r+1)b-t},$$

where r, t are arbitrary constants; all rational functions of given solutions that are again solutions can be constructed according to repetitions of the operations indicated in

$$kG(v), \quad G_1(v) + G_2(v), \quad (-1)^{(\eta+1)s}v^{sb/2}G_1(v)[G_2(v)]^s \qquad (\epsilon = 1, -1),$$

where k is an arbitrary constant, and $G(v)$, $G_1(v)$, $G_2(v)$ *are given solutions.*

COROLLARY 3. *Solutions of the type* $A(u)$ *exist only when* $b = 0$; $A(u)$ *is then as in Theorem* 6 *with* $\eta = 1$. *All solutions of the type* $B(v)$ *are obtained from Theorem* 6 *by taking* $\eta = 0$ *in* $B(v)$ *as there.*

The first part of this has some interesting consequences. Since $F(z) +$

$(-1)^s F(-z)$ is here the generator of an invariant polynomial sequence whose base ϕ is of index s, we have the formal expansion

$$F(z) + (-1)^s F(-z) = z^s \sum_{n=0}^{\infty} \phi_{2n+s} \frac{z^{2n}}{(2n+s)!},$$

since the left is changed into $(-1)^s$ times itself when z is replaced by $-z$, and $\phi_{2n+s+1}=0$. The characteristic is here $(-1, 0, (-1)^{n+s})$. If s is even, this becomes $(-1, 0, (-1)^n)$ which, as will be seen in §8, is the characteristic of each of the invariant polynomial sequences whose respective bases are E, L (cf. §1). Thus extensive tracts of the theories of the ϕ, E, L invariant polynomial sequences will be identical. One respect in which they may differ is more striking. By considering the special case in which the generator of ϕ is a rational function of z of the most general type possible, we easily find the following:

COROLLARY 4. *Let α, β be arbitrary constant integers, $(\alpha, \beta) \neq (0, 0)$, and the p_i, q_i $(i=0, 1, \cdots, \alpha;\ j=0, 1, \cdots, \beta)$ arbitrary constants; $(p_0, q_0) \neq (0, 0)$. Then ϕ defined by*

$$\phi_{2n+s+1} = 0, \qquad \sum_{j=0}^{\beta} q_j \frac{\phi_{2n+s-j}}{(2n+s-j)!} = \theta_n p_n,$$

where $\theta_n = 1$ or 0 according as $n \leqq \alpha$ or $n > \alpha$, is the base of an invariant polynomial sequence with index $(-1, 0, (-1)^{n+s})$.

Neither of E, L can be defined by a linear difference equation of constant order, since otherwise certain general circular functions would be algebraic.

Continuing with the general theorems we exhaust the possibilities in the next.

THEOREM 7. *All solutions of the functional equation of invariant generators of the types $A(u)\ R(u, v)$, $B(v)\ R(u, v)$, where $R(u, v)$ is any solution involving both u and v, and $A(u)$, $B(v)$ are functions of u alone, v alone respectively, are given by $A(u) = F(u) + F(-u)$, where $F(u)$ is an arbitrary function of u, and by $B(v) = H(v)$, where $H(v)$ is any solution of $H(v^{-1}) = H(v)$; the H equation has the elementary solution $H(v) \equiv k(v^r + v^{-r})$, where k, r are arbitrary constants, and $kH(u)$, $H_1(u) + H_2(u)$, $H_1(u)\ H_2(u)$, $H_1(u)/H_2(u)$, where $H(u)$, $H_1(u)$, $H_2(u)$ are given solutions, are further solutions.*

5. **Equivalent invariant polynomial sequences.** Applying §4 Theorem 3 to §2 Theorems 1, 2 we get the complete solution of the problem of equivalence for invariant polynomial sequences.

THEOREM 8. *Let the Appell polynomials $f_n(x)$, $h_n(x)$ have the respective bases λ, μ whose indices are s,t, and let the corresponding canonical polynomials be $F_n(x)$, $H_n(x)$, so that*

$$n!F_n(x) \equiv pf_n(x), \qquad n!\,H_n(x) \equiv qh_n(x),$$

where $p \neq 0, q \neq 0$ are arbitrary constants. Then

$$F_n'(x) = F_{n-1}(x), \qquad F_n(-x+b) = (-1)^{n+s}F_n(x),$$
$$H_n'(x) = H_{n-1}(x), \qquad H_n(-x+c) = (-1)^{n+t}H_n(x),$$
$$b \equiv -2\lambda_{s+1}/[(s+1)\lambda_s], \qquad c \equiv -2\mu_{t+1}/[(t+1)\mu_t].$$

Let i, j, m be arbitrary constant integers $\geqq 0$. Then

$$X_n(x) = G_n(x) \equiv F_{n+2i+s+m}\left(x + \frac{b-a}{2}\right),$$

$$X_n(x) = K_n(x) \equiv H_{n+2j+t+m}\left(x + \frac{c-a}{2}\right)$$

are solutions of

$$X_n'(x) = X_{n-1}(x), \quad X_n(-x+a) = (-1)^{n+m}X_n(x),$$

where a is an arbitrary constant, being equal to $-2\sigma_{m+1}/[(m+1)\sigma_m]$, where σ is the base of the general Appell polynomial in x and m is the index of σ.

Thus, according to the definition in §2, the G, K sequences are equivalent. Let us call the X sequence the *equalizing sequence* for G, K. The data in a specific application of Theorem 8 will be the f, h sequences. Without the theorem the invariant properties of these sequences must be investigated separately. The advantages of replacing f, h by F, H which are equivalent are obvious. By successive applications we get the following general result:

THEOREM 9. *By repeated applications of Theorem 8 to the equalizing sequences of pairs of sequences of Appell polynomials, and to an equalizing sequence and a sequence of canonical polynomials corresponding to a given sequence of Appell polynomials, any number of sequences of Appell polynomials can be transformed into the same number of equivalent sequences, all equalized with respect to one invariant polynomial sequence.*

An example is given in §8.

COROLLARY 5. *The equalizing sequence in Theorem 8 is unique, as also are the equivalent sequences which it equalizes, up to an arbitrary constant a in the argument of the equalizing sequence, and arbitrary constant integers $\geqq 0$ in the ranks of the polynomials equalized, and the arbitrary constant a in their arguments.*

6. **Rational invariant polynomial sequences.** If the generator $R(z, e^z)$ of ϕ is such that $R(u, v)$ is a rational function of the independent variables u, v, we shall call the invariant polynomial sequence with ϕ as base *rational*. The base of a rational invariant polynomial sequence will be called *rational*. The terms of a rational base are of course not necessarily rational numbers. The determination of all rational invariant polynomial sequences is reduced by the next theorem to that of all rational bases, which is done in §7. From §4 Theorem 3 we get the following:

THEOREM 10. *The set of all rational invariant polynomial sequences is completely and uniquely defined by the properties*

$$f'(x) = f_{n-1}(x), \quad f_n(-x + 2\phi_{s+1}/[(s+1)\phi_s]) = (-1)^{n+s} f_n(x),$$

where ϕ, whose index is s, is generated by $R(u, v)$, where $R(u, v)$ is any rational function of the independent variables u, v which is such that

$$v^b R(u,v) = (-1)^s R(-u, v^{-1}), \qquad b \equiv -2\phi_{s+1}/[(s+1)\phi_s],$$

and hence, for all such sequences, $-2\phi_{s+1}/[(s+1)\phi_s]$ is an integer.

7. **Rational bases.** Write the $R(u, v)$ of Theorem 10 in the form $R(u, v) \equiv v^c N(u, v)/D(u, v)$, where

$$N(u,v) \equiv \sum_{i=0}^{\alpha} v^i N(u), \qquad D(u,v) \equiv \sum_{j=0}^{\beta} v^j D(u)$$

are polynomials in u, v with highest common factor unity, the $N_i(u)$, $D_j(u)$ are polynomials in u alone of degree $\geqq 0$, and c is a constant integer $\gtreqless 0$. The trivial case in which $R(u, v)$ is a constant is excluded. The case $\alpha = \beta = c = 0$ was discussed in §4, Theorem 6, Corollary 4; it is included in the next theorem. As the solution of the pertinent functional equation presents no difficulty, we merely state the result, which can be verified by inspection.

THEOREM 11. *The general solution of the functional equation of rational generators,*

$$v^b R(u,v) = (-1)^s R(-u, v^{-1}),$$

in which b, s are arbitrary constant integers $\gtreqless 0$, and $R(u, v)$ is a rational function of the independent variables u, v is

$$R(u,v) = v^c N(u,v)/D(u,v),$$

where

c is an arbitrary constant intger $\gtreqless 0$;

$$N(u,v) = \sum_{i=0}^{\alpha}[v^iN_i(u) + \epsilon(-1)^{\eta\alpha}v^{\alpha-i}N_i(-u)],$$

$$D(u,v) = \sum_{j=0}^{\beta}[v^jD_j(u) + \epsilon(-1)^{(\eta+1)\beta}v^{\beta-j}D_j(-u)];$$

α, β are arbitrary constant integers such that

$$\alpha\beta \neq 0, \quad \alpha \geqq 0, \quad \beta \geqq 0, \quad \beta - \alpha = b + 2c;$$

no one of $N_0(u)$, $N_\alpha(u)$, $D_0(u)$, $D_\beta(u)$ is identically zero;

$N_i(u)$ $(i=0, \cdots, \alpha)$, $D_j(u)$ $(j=0, \cdots, \beta)$, with the exceptions just noted, are arbitrary polynomials in u alone of any degrees $\geqq 0$;

ϵ is a definite one of $1, -1$, and η a definite one of $0, 1$, and the same value of (ϵ, η) is to be used in both of $N(u,v)$, $D(u,v)$.

According to the values of (ϵ, η), rational invariant polynomial sequences fall into four mutually exclusive sets,each of which contains an infinity of sequences. The like holds when the numbers of the base are restricted to be rational, or to be in any given number field.

8. **The classic instances.** In the even suffix notation the sequences whose umbrae are B, G, E, L (cf. §1, end), have all terms, except B_1, G_1, of odd ranks, zero, and the signs alternate after the first term (rank 0).

The initial values necessary and sufficient for our purpose are

$$(B_0,B_1) = (1, -\tfrac{1}{2}), \qquad (G_0,G_1,G_2) = (0,1,-1),$$
$$(E_0,E_1) = (1,0), \qquad (L_0,L_1) = (\tfrac{1}{2},0);$$

hence the indices s of B, G, E, L are $0, 1, 0, 0$ respectively, and the corresponding Appell polynomials are

$$\beta_n(x) \equiv (x+B)^n, \qquad \gamma_n(x) \equiv (x+G)^n,$$
$$\eta_n(x) \equiv (x+E)^n, \qquad \lambda_n(x) \equiv (x+L)^n.$$

From the stated values of s and the first $s+2$ initial values in each case we find the values of b (§4, Theorem 3) for $\beta, \gamma, \eta, \lambda$ to be $1, 1, 0, 0$ respectively. Hence, by Theorem 3, we have

THEOREM 12. *The $\beta, \gamma, \eta, \lambda$ sequences of polynomials are solutions of the respective pairs of functional equations*

$$\beta_n'(x) = n\beta_{n-1}(x), \qquad \beta_n(-x+1) = (-1)^n\beta_n(x);$$
$$\gamma_n'(x) = n\gamma_{n-1}(x), \qquad \gamma_n(-x+1) = (-1)^{n+1}\gamma_n(x);$$
$$\eta_n'(x) = n\eta_{n-1}(x), \qquad \eta_n(-x) = (-1)^n\eta_n(x);$$
$$\lambda_n'(x) = n\lambda_{n-1}(x), \qquad \lambda_n(-x) = (-1)^n\lambda_n(x).$$

The known generators of B, G, E, L are (in the u, v form)

$$\frac{u}{v-1}, \quad \frac{2u}{v+1}, \quad \frac{2v}{v^2+1}, \quad \frac{uv}{v^2-1}$$

respectively, and these obviously verify §7, Theorem 11. [In particular, the integers α, β, b, c in Theorem 11 are here (there can be no confusion between these α, β and the umbrae)

$$(\alpha,\beta,b,c) = (0,1,1,0), \quad (0,1,1,0), \quad (0,2,0,1), \quad (0,2,0,1),$$

respectively ($c \equiv$ the exponent $\gtrless 0$ of the highest or the lowest power of v dividing the numerator; $\alpha \equiv$ the degree of the numerator in v after the division; $\beta \equiv$ the degree in v of the denominator; $b \equiv$ the integer already used in writing down the functional equations); the sign of c is so chosen that the denominator is not divisible by v].

To equalize the β, γ pair by §5 Theorem 8, take $\beta \equiv f$, $\gamma \equiv h$, and hence $(s, t) = (0, 1)$, $c = b = 1$, in the notation of §5. This gives

THEOREM 13. *If i, j, m are arbitrary constant integers $\geqq 0$, and p, q, a arbitrary constants, and if*

$$(n + 2i + m)!\,G_n(x) \equiv p\beta_{n+2i+m}\left(x + \frac{1-a}{2}\right),$$

$$(n + 2j + m + 1)!\,K_n(x) \equiv q\gamma_{n+2j+m+1}\left(x + \frac{1-a}{2}\right),$$

then $X_n(x) = G_n(x)$, $X_n(x) = K_n(x)$ are solutions of

$$X_n'(x) = X_{n-1}(x), \quad X_n(-x + a) = (-1)^{n+m}X_n(x).$$

In the same way we find for the η, λ pair

THEOREM 14. *If k, l, t are arbitrary constant integers $\geqq 0$, and r, g, b arbitrary constants, and if*

$$(n + 2k + t)!\,P_n(x) \equiv r\eta_{n+2k+t}\left(x - \frac{b}{2}\right),$$

$$(n + 2l + t)!\,Q_n(x) \equiv g\lambda_{n+2l+t}\left(x - \frac{b}{2}\right),$$

then $Y_n(x) = P_n(x)$, $Y_n(x) = Q_n(x)$ are solutions of

$$Y_n'(x) = Y_{n-1}(x), \quad Y_n(-x + b) = (-1)^{n+t}Y_n(x).$$

We now apply §5, Theorem 9. First equalize X, Y.

THEOREM 15. *If d, e, s are arbitrary constant integers $\geqq 0$, and w, h, c arbitrary constants, and if*

$$(n+2d+m+s)!S_n(x) \equiv wX_{n+2d+m+s}\left(x+\frac{a-c}{2}\right),$$

$$(n+2e+t+s)!T_n(x) \equiv hY_{n+2e+t+s}\left(x+\frac{b-c}{2}\right),$$

then $Z_n(x)=S_n(x)$, $Z_n(x)=T_n(x)$ are solutions of

$$Z_n'(x) = Z_{n-1}(x), \qquad Z_n(-x+c) = (-1)^{n+s}Z_n(x).$$

Replace $S_n(x)$ by its equivalent in terms of β, γ as given by combining the definitions in Theorems 13, 15, and similarly for $T_n(x)$, η, λ and Theorems 14, 15. Then finally we have the general equalization of $\beta, \gamma, \eta, \lambda$.

THEOREM 16. *If i, j, k, l, s are arbitrary constant integers $\geqq 0$, and p, q, r, g, c are arbitrary constants, and if*

$$(n+2i+s)!B_n(x) \equiv p\beta_{n+2i+s}\left(x+\frac{1-c}{2}\right),$$

$$(n+2j+s+1)!G_n(x) \equiv q\gamma_{n+2j+s+1}\left(x+\frac{1-c}{2}\right),$$

$$(n+2k+s)!E_n(x) \equiv r\eta_{n+2k+s}\left(x-\frac{c}{2}\right),$$

$$(n+2l+s)!L_n(x) \equiv g\lambda_{n+2l+s}\left(x-\frac{c}{2}\right),$$

then $W_n(x)=B_n(x)$, $W_n(x)=G_n(x)$, $W_n(x)=E_n(x)$, $W_n(x)=L_n(x)$ are solutions of

$$W_n'(x) = W_{n-1}(x), \qquad W_n(-x+c) = (-1)^{n+s}W_n(x).$$

To exhibit the particular form of this current for $B_n(x)$, $G_n(x)$ in the literature, we state the following

COROLLARY 6. *The equations*

$$U_n'(x) = U_{n-1}(x), \qquad U_n(-x-1) = (-1)^n U_n(x)$$

have the solutions

$$U_n(x) = \mathfrak{B}_n(x) \equiv \beta_n(x+1)/n!, \qquad U_n(x) = \mathfrak{E}_n(x) \equiv \eta_n(x+\tfrac{1}{2})/n!,$$

$$U_n(x) = \mathfrak{G}_n(x) \equiv \gamma_{n+1}(x+1)/(n+1)!, \qquad U_n(x) = \mathfrak{L}_n(x) \equiv \lambda_n(x+\tfrac{1}{2})/n!.$$

Glancing back over this section we see that everything in it, with the exception of the u, v forms of the generators, is an immediate consequence of the numerical values of the indices s and the first $s+2$ terms of the bases of the polynomials. That this should be so, and that the like holds also in the general case, is a remarkable simplification of the theory.

For those who may wish to pursue the β, γ, η, λ, and hence also the B, G, E, L, further by the methods of this paper, we add

$$2(1-2^n)B_n = G_n, \qquad (1-2^{n-1})B_n = L_n,$$

$$(2G)^n = 2n(E-1)^{n-1}, \qquad 2nE^{n-1} = (2G+1)^n,$$

all of which are well known and follow at once from trivial algebraic identities between the generators in their u, v forms.

9. **Remarks on notation and method.** Many writers on the Bernoulli and Euler numbers prefer a notation which makes the use of the symbolic method impossible, for example Nielsen in his *Traité Elémentaire des Nombres de Bernoulli* (Paris, 1923, pp. 9+398). His $(-1)^{n-1}B_n$, $n(-1)^n 2^{2(1-n)}T_n$ are our B_{2n}, G_{2n} $(n>0)$; his polynomials $2(n+1)!\,E_n(x)$, $n!B_n(x)$ are our $(x+G+1)^{n+1}$, $(x+B+1)^n$. By ignoring the well established symbolic method he is compelled (loc. cit., p. 46) to write his $E_n(x)$ in the form

$$\frac{1}{2}\frac{x^n}{n!} + \sum_{s=1}^{s\leq(n+1)/2} \frac{(-1)^{s-1}T_s x^{n-2s+1}}{(2s-1)!(n-2s+1)!2^{2s}},$$

which seems less suggestive and less tractable than its equivalent

$$(x+G+1)^{n+1}/[2(n+1)!].$$

As Neilsen in his preface emphasizes that the use of the functional equations is a "méthode élémentaire qui est beaucoup plus fondamentale que la méthode symbolique, développée notamment par Lucas" (*it was invented and very extensively applied to the Bernoulli and Euler numbers by J. Blissard fifteen years before Lucas' work was published*), it is well to point out what is indeed otherwise self-evident: *neither method is more fundamental than the other in any significant sense; they are abstractly identical.* For, the symbolic method, as we have shown, leads directly to the functional equations, and these are uniquely determined by the numerical values of s and the first $s+2$ terms of the respective bases, but not without them; conversely, the functional equations, together with the numerical values of s and the first $s+2$ terms of the bases, uniquely determine the generators, which are the fundamental formulas of the symbolic method in any given particular instances. That is, *each method implies and is implied by the other; they are*

thus formally equivalent in the sense of mathematical logic, or abstractly identical, as the term is used in algebra. To select a particular sequence given by the functional equations, the "elementary method" (to use Neilsen's name for it) adjoins a difference equation; the symbolic method presents the generator of the base, and again these procedures are abstractly identical.

Further, *the symbolic method, including the generators, is no more transcendental,* as has been carelessly alleged by certain writers, *than is the elementary.* For, the equality of generators is precisely matric equality, and this is exactly as transcendental as is mathematical induction, without which no formula inferred from the processes of the elementary method is proved, however obvious it may appear that the tedious induction will sustain the inference. Operations on generators are equivalent to the Cauchy addition, subtraction, multiplication and division of one-rowed matrices or, if preferred, of sequences, and these operations are abstractly identical with those of the elementary method. Heuristically, however, the advantage is with the symbolic method. This is abundantly evident on historical grounds, and is not affected by "elementary" reconstructions of theorems already known.

CALIFORNIA INSTITUTE OF TECHNOLOGY,
PASADENA, CALIFORNIA

ON THE DEGREE OF CONVERGENCE OF THE GRAM-CHARLIER SERIES*

BY

W. E. MILNE

The degree of convergence of Fourier series and of polynomial approximations has been extensively studied by Jackson,† and similar investigations have been made for series of Legendre polynomials,‡ for Sturm-Liouville series,§ for Laplace's series,‖ for Birkhoff's series,¶ and for other related problems.** In all these cases the interval over which the convergence applies is finite. It is therefore of interest to inquire what are the corresponding facts in a case where the interval is infinite. Such a case is afforded by the Gram-Charlier series, which, in addition to having an infinite interval, is of interest and importance as a mathematical statement of the law of errors†† and because of its application to the representation of frequency functions.

In the following pages we propose to examine the degree of convergence of the Gram-Charlier series and associated expansions, principally from the point of view of developments in characteristic solutions of homogeneous linear differential systems, a procedure which correlates the work closely with the previously cited papers on degree of convergence. The results obtained indicate that the Gram-Charlier series converges in general more slowly than the Fourier series for a similar function in a finite interval. Roughly speaking we may summarize the situation as follows: when the remainder after n terms of a Fourier series is $O(1/n^k)$ the remainder after n terms of the Gram-Charlier series is $O(1/n^{k/2})$.

1. If we denote the normal probability function by $\phi_0(x)$ and its derivative of order n by $\phi_n(x)$, then

$$\phi_0(x) = (2\pi)^{-1/2}e^{-x^2/2}, \qquad \phi_n(x) = H_n(x)\phi_0(x), \tag{1}$$

---

* Presented to the Society, March 30 and June 21, 1929; received by the editors March 7, 1929.

† Jackson, these Transactions, vol. 13 (1912), pp. 491–515.

‡ Jackson, these Transactions, vol. 13 (1912), pp. 305–318.

§ Jackson, these Transactions, vol. 15 (1914), pp. 439–466.

‖ Gronwall, these Transactions, vol. 15 (1914), pp. 1–30.

¶ Milne, these Transactions, vol. 19 (1918), pp. 143–156.

** Jackson, these Transactions, vol. 14 (1913), pp. 343–364, and vol. 22 (1921), pp. 158–166. See also Bulletin of the American Mathematical Society, vol. 27 (1921), pp. 415–431.

†† Charlier, Arkiv för Matematik, Astronomi, och Fysik, vol. 2 (1905).

where $H_n(x)$ is a polynomial of degree n, called an Hermite polynomial.* The functions $\phi_m(x)$ and $H_n(x)$ are biorthogonal in the infinite interval, so that

$$\int_{-\infty}^{\infty} \phi_m(x)H_n(x)dx = \begin{cases} 0 & \text{if } m \neq n, \\ n! & \text{if } m = n. \end{cases} \tag{2}$$

Hence we have the formal expansion of an arbitrary function $\phi(x)$

$$\phi(x) = a_0\phi_0(x) + a_1\phi_1(x) + a_2\phi_2(x) + \cdots \tag{3}$$

in which the coefficients are determined by the formulas

$$a_n = (1/n!)\int_{-\infty}^{\infty} \phi(x)H_n(x)dx. \tag{4}$$

The series (3) is the Gram-Charlier series. The convergence of this and related series has been investigated by Myller-Lebedeff,† Watson,‡ Cramér,§ Szegö,‖ Rotach,¶ Hille,** Stone,†† and others.

The functions $\phi_n(x)$ and $H_n(x)$ respectively satisfy the adjoint differential equations

$$u'' + xu' + (\lambda + 1)u = 0, \tag{5}$$

and

$$v'' - xv' + \lambda v = 0, \tag{6}$$

for integral values of λ, $\lambda = n$. If v is any solution of (6) the function

$$u = e^{-x^2/2}v$$

is a solution (5) and conversely. The transformations

$$u = e^{-x^2/4}w, \qquad v = e^{x^2/4}w$$

carry (5) and (6) respectively into the self-adjoint equation‡‡

* Hermite, Comptes Rendues, vol. 58, p. 93.

† Myller-Lebedeff, Mathematische Annalen, vol. 64 (1907), p. 388.

‡ Watson, Proceedings of the London Mathematical Society, (2), vol. 8 (1910), pp. 393–421.

§ Cramér, Den Sjette Skandinaviske Matematikerkongres i København, Kongresberetningen, Copenhagen, 1926, pp. 399–425.

‖ Szegö, Mathematische Zeitschrift, vol. 25 (1926), pp. 112–115.

¶ Rotach, Promotionsarbeit, Zurich, Ceuf, 1925, 33 pp.

** Hille, Annals of Mathematics, (2), vol. 27 (1926), pp. 427–464.

†† Stone, Annals of Mathematics, (2), vol. 29 (1928), pp. 1–11.

‡‡ Equation (7) is sometimes called Weber's equation, and is satisfied by the harmonic functions associated with the parabolic cylinder. See Weber, Mathematische Annalen, vol. 1, p. 1; Watson, loc. cit.; Whittaker, Proceedings of the London Mathematical Society, vol. 35 (1902), p. 417.

$$w'' + (\lambda + 1/2 - x^2/4)w = 0. \tag{7}$$

Corresponding to the three equations (5), (6), and (7) there are three distinct expansions in terms of characteristic functions, each one of which can be transformed into any other by means of the above relationships connecting u, v, and w.

2. Instead of dealing directly with the series (3) it is convenient to use as the basis of the following investigation the set of normalized orthogonal solutions of (7),

$$W_0(x), W_1(x), W_2(x), \cdots,$$

defined by the equation

$$W_n(x) = (2\pi)^{-1/4}(n!)^{-1/2}e^{-x^2/4}H_n(x). \tag{8}$$

From the well known properties of Hermite polynomials we can easily write down a number of important formulas as follows:

$$(n+1)^{1/2}W_{n+1}(x) + xW_n(x) + n^{1/2}W_{n-1}(x) = 0; \tag{9}$$

$$W_n'(x) = -(x/2)W_n(x) - n^{1/2}W_{n-1}(x); \tag{10}$$

$$W_n'(x) = (1/2)[(n+1)^{1/2}W_{n+1}(x) - n^{1/2}W_{n-1}(x)]; \tag{11}$$

$$W_n''(x) + (n + 1/2 - x^2/4)W_n(x) = 0; \tag{12}$$

$$\sum_{m=0}^{n-1} W_m(x)W_m(s) = n^{1/2}\{W_{n-1}(x)W_n(s) - W_n(x)W_{n-1}(s)\}(x-s)^{-1}; \tag{13}$$

$$\int_{-\infty}^{\infty} W_m(x)W_n(x)dx = \begin{cases} 1 & \text{if } m = n, \\ 0 & \text{if } m \neq n. \end{cases} \tag{14}$$

From equation (11),

$$W_{n+1}(x) = (n+1)^{-1/2}\{2W_n'(x) + n^{1/2}W_{n-1}(x)\}.$$

If we form a similar equation for $W_{n-1}(x)$ and substitute its value into the above expression for $W_{n+1}(x)$, then substitute again for $W_{n-3}(x)$, and continue the process the following two formulas are obtained (in which we have changed subscripts from $n+1$ to n), the first for the case in which n is even and the second for the case in which n is odd:

$$\begin{aligned} W_n(x) = 2[&n^{-1/2}W_{n-1}'(x) + \{(n-1)/(n(n-2))\}^{1/2}W_{n-3}'(x) + \cdots \\ &+ \{(n-1)\cdots 7\cdot 5\cdot 3/(n\cdots 6\cdot 4\cdot 2)\}^{1/2}W_1'(x)] \\ &+ \{(n-1)\cdots 7\cdot 5\cdot 3\cdot 1/(n\cdots 6\cdot 4\cdot 2)\}^{1/2}e^{-x^2/4}(2\pi)^{-1/4}; \end{aligned} \tag{15}$$

$$\begin{aligned} W_n(x) = 2[&n^{-1/2}W_{n-1}'(x) + \{(n-1)/(n(n-2))\}^{1/2}W_{n-3}'(x) + \cdots \\ &+ \{(n-1)\cdots 6\cdot 4\cdot 2/(n\cdots 5\cdot 3\cdot 1)\}^{1/2}W_0'(x)]. \end{aligned} \tag{16}$$

Since the W's with odd subscripts vanish at 0 and ∞ we find from (15)

$$(17)\qquad \int_0^\infty W_n(x)dx = (\pi/2)^{1/4}\{1\cdot 3\cdot 5\cdots(n-1)/(2\cdot 4\cdot 6\cdots n)\}^{1/2},$$

when n is even. On the other hand when n is odd (16) may be written

$$\begin{aligned}(18)\qquad W_n(x) &= 2\{2\cdot 4\cdot 6\cdots(n-1)/(1\cdot 3\cdot 5\cdots n)\}^{1/2}[W_0'(x) + (1/2)^{1/2}W_2'(x)\\ &\quad + (1\cdot 3/(2\cdot 4))^{1/2}W_4'(x) + \cdots\\ &\quad + \{1\cdot 3\cdot 5\cdots(n-2)/(2\cdot 4\cdot 6\cdots(n-1))\}^{1/2}W_{n-1}'(x)].\end{aligned}$$

When m is even, $W_m(x)$ vanishes at ∞, but at $x=0$ we have

$$(19)\qquad W_m(0) = (-1)^{m/2}(2\pi)^{-1/4}(1\cdot 3\cdot 5\cdots(m-1)/(2\cdot 4\cdot 6\cdots m))^{1/2},$$

so that

$$\begin{aligned}(20)\qquad \int_0^\infty W_n(x)dx &= -2(2\pi)^{-1/4}[2\cdot 4\cdot 6\cdots(n-1)/(1\cdot 3\cdot 5\cdots n)]^{1/2}\\ &\quad\times[1 - 1/2 + 1\cdot 3/(2\cdot 4) - 1\cdot 3\cdot 5/(2\cdot 4\cdot 6) + \cdots\\ &\quad\pm 1\cdot 3\cdot 5\cdots(n-2)/(2\cdot 4\cdot 6\cdots(n-1))]\end{aligned}$$

when n is odd. The series in brackets is the series, up to terms of degree $n-1$, for the expansion of $(1+x^2)^{-1/2}$ at the value $x=1$, and therefore the sum of the terms in brackets in (20) differs from $2^{-1/2}$ by a quantity less than the last term. Stirling's formula for $n!$ enables us to estimate the magnitude of this last term, which we find to be $(2/(\pi n))^{1/2}(1+O(n^{-1}))$. Therefore (20) may be rewritten

$$(21)\qquad \int_0^\infty W_n(x)dx = -(2/\pi)^{1/4}[2\cdot 4\cdot 6\cdots(n-1)/(1\cdot 3\cdot 5\cdots n)]^{1/2}(1+\epsilon),$$

when n is odd.

Finally if (17) and (21) are evaluated by means of Stirling's formula we have

$$(22)\qquad \int_0^\infty W_n(x)dx = \pm n^{-1/4}(1+\epsilon_n),\quad +\text{ if } n \text{ is even},\quad -\text{ if } n \text{ is odd}.$$

3. To make clear the subsequent discussion it is needful to describe in some detail the character of the function $W_n(x)$. First of all $W_n(x)$ is even when n is even and odd when n is odd, so that we may limit the discussion to the case where x is positive. By well known theorems of oscillation we see from the differential equation (7) that in the interval $0\leqq x<h$ (where for brevity we let $h=(4n+2)^{1/2}$) the function $W_n(x)$ oscillates with increasing amplitudes, increasing intervals between the roots, and with decreasing slopes

at the roots. In the interval $h<x<\infty$, $W_n(x)$ does not oscillate, but approaches zero to an infinite order as x becomes infinite. The roots of $W_n(x)$ are identical with the roots of $H_n(x)$, which are known to be all real, n in number, symmetrically placed with respect to the origin, and all in the interval $-h<x<h$. The largest maximum of $W_n(x)$ occurs in the interval $r<x<h$, where r denotes the largest root. The approximate location of this largest root may be estimated as follows. In (7) we make the substitution

$$x = h - (2/h)^{1/3}t,$$

which reduces it to

$$d^2w/dt^2 + t[1 - (2/h)^{4/3}t/4]w = 0. \tag{23}$$

We now form two comparison equations

$$d^2u/dt^2 + tu = 0,$$

and

$$d^2v/dt^2 + (t/2)v = 0.$$

We solve (23) with those conditions at $t=0$ which furnish the solution $W_n(x)$ and solve the two comparison equations with the same initial values. We denote the first positive roots of w, u, and v by t_2, t_1, and t_3 respectively. Then if n is large enough that

$$1 - (2/h)^{4/3}t_3/4 > 1/2$$

we see by comparing the differential equations that

$$t_1 < t_2 < t_3.$$

Now t_1 and t_3 are not entirely independent of n because of the initial conditions. But at $t=0$, u and du/dt have opposite signs, from which we may conclude that t_1 has a lower bound c_1, not zero, and independent of n. By numerical calculation we find from the equation $d^2u/dt^2+tu=0$ that it is possible to take $c_1=7^{1/3}=1.91$. Also t_3 has a finite upper bound c_3, independent of n. Consequently

$$h - (2/h)^{1/3}c_3 < r < h - (2/h)^{1/3}c_1. \tag{24}$$

To the function $W_n(x)$ may be applied certain inequalities based on Hille's investigations* in the case of Hermitian polynomials. Following his methods we are able to show that

$$|W_n(x)| < K(h^2 - x^2)^{-1/4}, \quad -r \leqq x \leqq r, \tag{25}$$

where K is a constant independent of n. We have also

$$|W_n(x)| < K_1 n^{-1/12} \tag{26}$$

* Hille, loc. cit., pp. 433–436.

for all values of x. A further inequality applicable to the interval $h<x<\infty$ may be established by the following considerations: When n is even, the function $W_n(x)$ is positive and decreasing and its curve is concave upward for all values of x greater than h. If we denote by $w_0, w_1, w_2, \cdots, w_0', w_1', w_2', \cdots$ the values of $W_n(x)$ and $W_n'(x)$ at the points $h, h+1, h+2, \cdots$, the following inequalities are apparent:

$$w_{k-1} > -w_k'. \tag{27}$$

From the equation (7) we see that, in the interval from $h+k$ to $h+k+1$,

$$w'' > \tfrac{1}{4}[(h+k)^2 - h^2]w_{k+1}.$$

After integrating from $h+k$ to $h+k+1$ and dropping some terms to strengthen the inequality we have

$$-w_k' > \tfrac{1}{2}khw_{k+1}. \tag{28}$$

From (27) and (28) it follows that

$$w_{k+1} < 2w_{k-1}/(kh)$$

and from this we get by successive substitutions

$$w_2 < 2w_0/h, \quad w_4 < 4w_0/(3h^2), \quad w_6 < 8w_0/(3\cdot 5h^3), \cdots.$$

Therefore we may conclude that for $x \geqq h+2k$

$$|W_n(x)| < K_2 n^{-k/2}. \tag{29}$$

The case in which n is odd is similarly treated.

4. In addition to estimates of the magnitude of $W_n(x)$ we shall require estimates of the magnitude of the integral of $W_n(x)$. For this purpose let $r_1, r_2, r_3, \cdots, r$ denote the positive roots of $W_n(x)$. We then have the inequality

$$\left|\int_{r_{i-1}}^{r_i} W_n(x)dx\right| < \left|\int_{r_i}^{r_{i+1}} W_n(x)dx\right|. \tag{30}$$

Let

$$m_i = \pi/(r_{i+1} - r_i).$$

Then from (12) we have

$$(h^2 - r_{i+1}^2)^{1/2} < 2m_i < (h^2 - r_i^2)^{1/2},$$

as may be seen by comparing the intervals between successive roots of those solutions of the differential equations

$$d^2u/dx^2 + \tfrac{1}{4}(h^2 - r_{i+1}^2)u = 0,$$
$$d^2u/dx^2 + \tfrac{1}{4}(h^2 - x^2)u = 0,$$
$$d^2u/dx^2 + \tfrac{1}{4}(h^2 - r_i^2)u = 0,$$

which vanish at $x = r_i$.

Hence there is a value x_i of x between r_i and r_{i+1} such that

$$(h^2 - x_i^2)^{1/2} = 2m_i.$$

The equation (12) may now be written in the form

$$W_n''(x) + m_i^2 W_n(x) = \tfrac{1}{4}(x^2 - x_i^2)W_n(x),$$

from which we obtain the integral equation

$$W_n(x) = (1/m_i)W_n'(r_{i+1}) \sin m_i(x - r_{i+1}) + \frac{1}{4m_i}\int_{r_{i+1}}^{x} \{\sin m_i(x - s)\}(s^2 - x_i^2)W_n(s)ds. \tag{31}$$

The integral of this from r_i to r_{i+1} is

$$\int_{r_i}^{r_{i+1}} W_n(x)dx = -(1/m_i)^2 2W_n'(r_{i+1}) + \frac{1}{4m_i}\int_{r_i}^{r_{i+1}} dx \int_{r_{i+1}}^{x} \{\sin m_i(x - s)\}(s^2 - x_i^2)W_n(s)ds.$$

Now

$$|s^2 - x_i^2| < r_{i+1}^2 - r_i^2 < 2\pi r_{i+1}/m_i < 4\pi r_{i+1}(h^2 - r_{i+1}^2)^{-1/2},$$

and from (25)

$$|W_n(s)| < K(h^2 - r_{i+1}^2)^{-1/4}$$

so that using these inequalities together with the inequalities for m_i we have

$$\left|\frac{1}{4m_i}\int_{r_i}^{r_{i+1}} dx \int_{r_{i+1}}^{x} \{\sin m_i(x - s)\}(s^2 - x_i^2)W_n(s)ds\right| < C_2 r_{i+1}(h^2 - r_{i+1}^2)^{-9/4}, \tag{32}$$

where C_2 is a constant independent of n.

It will be found upon reference to Hille's arguments from which (25) was derived that we also have the inequality

$$|W_n'(x)| < K'(h^2 - x^2)^{1/4}, \tag{33}$$

where K' is independent of n, and x is between $-r$ and r. Then using (33) in connection with (32) we are led to the inequality

$$\left|\int_{r_i}^{r_{i+1}} W_n(x)dx\right| < 8K'(h^2 - r_{i+1}^2)^{-3/4} + C_2 r_{i+1}(h^2 - r_{i+1}^2)^{-9/4}. \tag{34}$$

From the fact that the successive integrals satisfy the inequalities (30) and alternate in sign we see at once that

$$(35)\qquad \left| \int_0^x W_n(x)\,dx \right| < 8K'(h^2 - r_i^2)^{-3/4} + C_2 r_i (h^2 - r_i^2)^{-9/4},\ x \leqq r_i.$$

In particular if x is in the interval $-N \leqq x \leqq N$, where $N = (3n+2)^{1/2}$,

$$(36)\qquad \left| \int_0^x W_n(x)\,dx \right| < 8K' n^{-3/4} + O(n^{-7/4}).$$

Since the extrema of the integral occur only at the roots of $W_n(x)$ its largest maximum absolute value must be found either at one of the roots or else at $x = \infty$. If the former is true we find from (35) together with the fact that the largest root is limited by (24) that

$$(37)\qquad \left| \int_0^x W_n(x)\,dx \right| < 2K' n^{-1/4}[1 + \epsilon],$$

for all values of x. On the other hand if the largest maximum occurs at $x = \infty$ its value is given by (22).

5. We are now in a position to investigate questions concerning degree of convergence. First of all it will be interesting and instructive to consider a series formed for a finite interval and to watch what happens as the ends of the interval recede to $-\infty$ and ∞. Together with equation (7) let us consider a pair of Sturmian boundary conditions at the ends of a finite interval $(-c, c)$, and let $w_n(x)$ denote the normalized orthogonal characteristic solution of this system corresponding to the $(n+1)$th characteristic number λ_n. The formal expansion of an arbitrary function $f(x)$ will be

$$(38)\qquad f(x) = c_0 w_0(x) + c_1 w_1(x) + c_2 w_2(x) + c_3 w_3(x) + \cdots.$$

This is a Sturm-Liouville series, so that the results obtained by Jackson regarding the degree of convergence of these series may be applied directly. We are told for example that if $f(x)$ has a continuous kth derivative of bounded variation in the interval $(-c, c)$, and if $f(x)$ and its first $k-1$ derivatives vanish at $-c$ and c, then the first n terms of the series (38) represents $f(x)$ with an error which is $O(\lambda_n^{-k/2})$.

When c is allowed to become infinite the characteristic number λ_{n+1} approaches n as a limit and each term of the series (38) approaches* the corresponding term of the series

* Cf. Weyl, Göttinger Nachrichten, 1910, p. 442; also Milne, these Transactions, vol. 30 (1928), pp. 797–802.

$$f(x) = C_0W_0(x) + C_1W_1(x) + C_2W_2(x) + C_3W_3(x) + \cdots \tag{39}$$

where

$$C_n = \int_{-\infty}^{\infty} f(x)W_n(x)dx. \tag{40}$$

We are thus led to conjecture that if $f(x)$ has a kth derivative of bounded variation in the infinite interval, and if $f(x)$ and its derivatives vanish at $\pm\infty$, then the first n terms of (39) will represent $f(x)$ with an error which is $O(n^{-k/2})$. The essential difference between this result and that for the finite interval lies in the distribution of the characteristic numbers in the two cases. For the finite interval the nth characteristic number is of the order of magnitude of $(\pi n/c)^2$, while for the infinite interval $\lambda_{n+1}=n$. Thus in the former case the remainder term is $O(n^{-k})$ while in the latter we conjecture that it is $O(n^{-k/2})$.

Subsequent theorems show that this conjecture is substantially correct.

6. Instead of attempting to carry out rigorously the line of thought suggested in the preceding paragraph we shall deal directly with the series (39). Letting $S_n(x)$ denote the sum of the first n terms of this series we may state the first theorem as follows:

THEOREM I. *If $f(x)$ has a continuous kth derivative of bounded variation, and if $xf^{(k)}(x)$, $x^2f^{(k-1)}(x)$, $\cdots$, $x^{k+1}f(x)$ also are of bounded variation in the infinite interval, then*

$$S_n(x) = f(x) + O(n^{-k/2}), \quad -N \leqq x \leqq N,$$

and

$$S_n(x) = f(x) + O(n^{-(3k-1)/6}), \quad -\infty < x < \infty.$$

These relations hold uniformly with respect to x in the specified intervals.

From (8) and (40) we may set

$$C_n = (2\pi)^{-1/4}(n!)^{-1/2}\int_{-\infty}^{\infty} e^{-x^2/4}f(x)H_n(x)dx,$$

whence integrating by parts k times with the aid of the relation

$$H_n(x)dx = -(n+1)^{-1}\,dH_{n+1}(x),$$

and observing that the integrated terms vanish at the limits, we obtain

$$C_n = [(n+1)(n+2)\cdots(n+k)]^{-1/2}\int_{-\infty}^{\infty} G_k(x)W_{n+k}(x)dx,$$

in which

$$G_k(x) = e^{x^2/4}\frac{d^k}{dx^k}[e^{-x^2/4}f(x)].$$

From the hypotheses it is clear first of all that $xG_k(x)$ is of bounded variation in the infinite interval, and since $f^{(k)}(x)$ is also of bounded variation it is easy to show that $G_k(x)$ is of bounded variation in the infinite interval. Hence we may set

$$G_k(x) = g_1(x) - g_2(x)$$

where $g_1(x)$ and $g_2(x)$ are positive or zero, continuous, monotone increasing, and bounded. By the second law of the mean

$$\int_{-N}^{N} G_k(x)W_{n+k}(x)dx = g_1(N)\int_{\xi_1}^{N} W_{n+k}(x)dx - g_2(N)\int_{\xi}^{N} W_{n+k}(x)dx$$

where ξ_1 and ξ_2 are in the interval $(-N, N)$. Therefore by (36)

$$\int_{-N}^{N} G_k(x)W_{n+k}(x)dx = O(n^{-3/4}).$$

Next we may write

$$\int_{N}^{\infty} G_k(x)W_{n+k}(x)dx = \int_{N}^{\infty} [xG_k(x)][W_{n+k}(x)/x]dx.$$

Since $xG_k(x)$ is of bounded variation, and since

$$\int_{\xi}^{\infty} [W_{n+k}(x)/x]dx = -(1/\xi)\int_{0}^{\xi} W_{n+k}(x)dx + \int_{\xi}^{\infty} x^{-2}dx\int_{0}^{x} W_{n+k}(x)dx = O(n^{-3/4})$$

if $\xi > N$, we may deduce in the same manner as above that

$$\int_{N}^{\infty} G_k(x)W_{n+k}(x)dx = O(n^{-3/4}).$$

The same is true of the integral from $-\infty$ to $-N$, so that finally

$$\int_{-\infty}^{\infty} G_k(x)W_{n+k}(x)dx = O(n^{-3/4})$$

and

$$C_n = O(n^{-k/2-3/4}).$$

Therefore in the interval $(-N, N)$ the inequality (25) gives

$$C_n W_n(x) = O(n^{-k/2-1}),$$

and hence in the same interval

$$\sum_{m=n}^{\infty} C_m W_m(x) = O(n^{-k/2}).$$

In view of the fact that under the given hypotheses the series is known to converge to $f(x)$ the first conclusion stated by the theorem follows. To establish the second we have only to use the inequality (26) in place of (25).

In the statement and proof of the foregoing theorem continuity of the kth derivative has been assumed. An examination of the methods of proof will reveal that this restriction can be to some extent dispensed with. For example the results can still be shown to hold true even if $f^{(k)}(x)$ has a finite number of finite discontinuities.

It may also be observed that the other conditions of the hypotheses of the theorem can be variously stated. Any set of conditions that will insure the convergence of the series to $f(x)$ and at the same time make the functions $G_k(x)$ and $xG_k(x)$ of limited variation in the infinite interval will do.

From Theorem I we may derive directly a corresponding theorem applicable to the Gram-Charlier series (3).

THEOREM II. *If $\phi(x)$ has a continuous kth derivative of bounded variation, and if*

$$xe^{x^2/4}\phi^{(k)}(x), \quad x^2e^{x^2/4}\phi^{(k-1)}(x), \cdots, x^{k+1}e^{x^2/4}\phi(x)$$

are of bounded variation in the infinite interval, then

$$|\Sigma_n(x) - \phi(x)| < Mn^{-k/2}(1 + x^2)^{1/6}e^{-x^2/4},$$

where $\Sigma_n(x)$ is the sum of n terms of (3) and M is a constant independent of n.

If $f(x) = e^{x^2/4}\phi(x)$ then $f(x)$ satisfies the hypotheses of Theorem I, and if we multiply (39) by $e^{-x^2/4}$ we obtain (3). From Theorem I we see that the remainder after n terms of (3) will be

$$O(n^{-k/2})e^{-x^2/4}$$

in the interval $-N \leqq x \leqq N$ and will be

$$O(n^{-k/2+1/6})e^{-x^2/4}$$

outside this interval. Now when $|x| > N$ it is clear that

$$n^{1/6} < (1 + x^2)^{1/6}$$

whence the truth of Theorem II is apparent.

Since the function $(1+x^2)^{1/6}e^{-x^2/4}$ is bounded we have

COROLLARY I. *With the hypotheses of Theorem* II

$$\sum_n(x) = \phi(x) + O(n^{-k/2}), \quad -\infty < x < \infty.$$

It will be noted that Theorem I requires $x^{k+1}f(x)$ to be of bounded variation in the infinite interval, which carries with it the implication that for x large

$$f(x) = O(x^{-k-1}).$$

This condition can be somewhat lightened, at the expense however of a corresponding increase in the remainder term for large values of x. We suppose that $f(x)$ satisfies the hypotheses of Theorem I, let

$$f(x) = S_n(x) + R_n(x),$$

multiply this by $(1+x^2)$ and get rid of the x^2 in the terms of the series by two successive applications of the formula

$$xW_p(x) = -(p+1)^{1/2}W_{p+1}(x) - p^{1/2}W_{p-1}(x).$$

Then we have

$$f_1(x) = C_0'W_0(x) + C_1'W_1(x) + \cdots + C_{n-2}'W_{n-2}(x) + R_{1,n}(x),$$

in which

$$f_1(x) = (1+x^2)f(x),$$

$$C_p' = [(p-1)p]^{1/2}C_{p-2} + (2p+2)C_p + [(p+1)(p+2)]^{1/2}C_{p+2},$$

and

$$\begin{aligned} R_{1,n}(x) = {} & \{[(n-2)(n-1)]^{1/2}C_{n-3} + 2nC_{n-1}\}W_{n-1}(x) \\ & + \{[(n-1)n]^{1/2}C_{n-2} + (2n+2)C_n\}W_n(x) \\ & + \{[n(n+1)]^{1/2}C_{n-1}\}W_{n+1}(x) \\ & + \{[(n+1)(n+2)]^{1/2}C_n\}W_{n+2}(x) \\ & + R_n(x)(1+x^2). \end{aligned}$$

In the course of the proof of Theorem I we have shown that

$$C_n = O(n^{-k/2-3/4})$$

and this fact together with (25), (26), and the inequality

$$n^{1/6} < (1+x^2)^{1/6}, \qquad |x| > N,$$

enables us to show that

$$R_{1,n}(x) = O(n^{-k/2})(1+x^2)^{7/6}.$$

Use of the recursion formula shows that the coefficients which we have designated by C_p' are actually the coefficients which we would have obtained

in the formal expansion of the function $f_1(x)$ itself, and therefore $R_{1,n}(x)$ is the remainder after n terms of the series (39) formed for the function $f_1(x)$. Therefore we have the following modification of Theorem I.

THEOREM III. *If $f(x)$ has a continuous kth derivative of bounded variation, and if $x^{k-1}f(x)$, $x^{k-2}f'(x)$, $x^{k-3}f''(x)$, $\cdots$, $x^{-1}f^{(k)}(x)$ are also of bounded variation when x is large, then*

$$S_n(x) = f(x) + O(n^{-k/2})(1 + x^2)^{7/6},$$

for all values of x.

It is this theorem which is to be regarded as closest analogue to the theorem of Jackson on Sturm-Liouville series referred to above.

The statement and proof of the corresponding theorem for the case of the series (3) may be left to the reader.

7. Since the results obtained in the preceding paragraph indicate a rate of convergence much slower than in the corresponding situation for Fourier series the question naturally arises whether these conclusions are inevitable or are merely due to inadequate methods of proof. The following example shows that no substantial improvement in the conclusions stated by the theorems need be expected.

Let $f(x) = |x|e^{-x^2/4}$. This function satisfies the hypotheses of Theorem I for $k=1$ except that the derivative has a finite discontinuity at $x=0$. But we have seen that the proof of Theorem I can still be carried through in spite of a finite number of discontinuities of the derivative. We multiply the identity (13) by the function $|s|e^{-s^2/4}$, integrate with respect to s from $-\infty$ to ∞, assume that is odd, then set $x=0$, and have

$$S_n(0) = n^{1/2}W_{n-1}(0)\left[\int_{-\infty}^{0} - \int_{0}^{\infty}\right] e^{-s^2/4}W_n(s)ds.$$

From equations (1) and (8)

$$n^{1/2}\left[\int_{-\infty}^{0} - \int_{0}^{\infty}\right] e^{-s^2/4}W_n(s)ds = 2W_{n-1}(0)$$

so that

$$S_n(0) = 2W_{n-1}^2(0) = 2(2\pi)^{-1/2}[1\cdot 3\cdot 5\cdots(n-2)/(2\cdot 4\cdot 6\cdots(n-1))].$$

The use of Stirling's formula shows that this last expression is equal to

$$(2/\pi)n^{-1/2} + O(n^{-3/2}).$$

Since $f(0)=0$

$$|S_n(0) - f(0)| > (1/\pi)n^{-1/2}.$$

This example therefore shows that a stronger conclusion than the one stated for Theorem I is not to be expected.

As a second example designed to show that the hypotheses about vanishing at infinity cannot be dispensed with, consider the function $f(x)=1$, which satisfies all the conditions of Theorem I for any k except that of vanishing at infinity. We note that $C_n=0$ if n is odd, and C_n is twice the expression given by (17) if n is even. Therefore if we integrate (18) from $-\infty$ to x and compare coefficients with the above values of C_n we find that

$$S_n(x) = (\pi/2)^{1/4}[1\cdot 3\cdot 5\cdots n/(2\cdot 4\cdot 6\cdots (n-1))]^{1/2}\int_{-\infty}^{x} W_n(x)dx,$$

or

$$S_n(x) = n^{1/4}(1+\epsilon_n)\int_{-\infty}^{x} W_n(x)dx,$$

where n is odd. From this it is plain that

$$\lim_{x=\pm\infty} S_n(x) = 0,$$

so that for any given n we may always choose x so large that

$$f(x) - S_n(x) > 1-\epsilon, \quad \epsilon > 0.$$

Thus no relation of the form

$$|f(x) - S_n(x)| = O(n^{-1/2})$$

can hold uniformly in the infinite interval.

We may remark however that by means of (36) and (20) it may be shown that

$$(41) \qquad S_n(x) = n^{1/4}(1+\epsilon_n)\left[\int_{-\infty}^{0} W_n(x)dx + \int_{0}^{x} W_n(x)dx\right]$$

$$= 1 + O(n^{-1/2}) + n^{1/4}(1+\epsilon_n)\int_{0}^{x} W_n(x)dx$$

$$= 1 + O(n^{-1/2}), \quad -N \leqq x \leqq N,$$

$$= 1 + O(1), \quad -\infty < x < \infty.$$

8. Theorems I, II, and III have stated sufficient conditions for a given degree of convergence. The next two will furnish certain necessary conditions. The first of these two theorems is closely analogous to one given by Jackson* for the case of Fourier series and shows that the existence of deriva-

* Jackson, *Über die Genauigkeit der Annäherung stetiger Funktionen durch rationale Funktionen gegebenen Grades und trigonometrische Summen gegebener Ordnung*, Dissertation, Göttingen, 1911; Satz XVIII.

tives of certain orders is necessary for a given degree of convergence.

THEOREM IV. *If a series*

$$S(x) = A_0W_0(x) + A_1W_1(x) + A_2W_2(x) + \cdots$$

converges so that for any n its remainder after n terms is

$$R_n(x) + O(n^{-k-\epsilon}), \quad k \text{ an integer}, \quad \epsilon > 0,$$

uniformly for x in any interval, then the sum S(x) has a continuous derivative of order 2k in the same interval.

Consider the sum

$$\sum_{m=n}^{n+p} m^q A_m W_m(x) = \sum_{m=n}^{n+p} m^q [R_m(x) - R_{m+1}(x)].$$

By partial summation this may be put in the form

$$\sum_{m=n}^{n+p+1} [(m+1)^q - m^q]R_{m+1}(x) + n^q R_n(x) - (n+p)^q R_{n+p+1}(x).$$

Now

$$(m+1)^q - m^q < q(m+1)^{q-1}.$$

Hence the general term under the sign of summation is less in absolute value than

$$Mq(m+1)^{q-1-k-\epsilon}, \qquad M \text{ a constant},$$

from which we see that the sum itself is

$$O(n^{q-k-\epsilon}).$$

The two terms not under the summation sign are clearly also

$$O(n^{q-k-\epsilon})$$

so that

$$\sum_{m=n}^{n+p} m^q A_m W_m(x) = O(n^{q-k-\epsilon}).$$

From this we conclude that the set of series

$$\sum_{m=0}^{\infty} m^q A_m W_m(x) \qquad (q = 1, 2, \cdots, k)$$

all converge uniformly in the given interval.

Now from equation (12)

$$\sum_{m=0}^{\infty} A_m W_m''(x) = -\sum_{m=0}^{\infty} A_m(m + 1/2 - x^2/4)W_m(x)$$

$$= -\sum_{m=0}^{\infty} mA_m W_m(x) + (x^2/4 - 1/2)\sum_{m=0}^{\infty} A_m W_m(x).$$

From the uniform convergence of the two series on the right follows the uniform convergence of the series of second derivatives, which in turn establishes the existence and continuity of the second derivative of $S(x)$. To proceed, we differentiate twice the relation

$$S''(x) = (x^2/4 - \tfrac{1}{2})S(x) - \sum_{n=0}^{\infty} nA_n W_n(x),$$

substitute for $W_n''(x)$ from (12), and obtain formally

$$S^{(4)}(x) = (x^2/4 - \tfrac{1}{2})S''(x) + xS'(x) + \tfrac{1}{2}S(x)$$

$$+ (\tfrac{1}{2} - x^2/4)\sum_{n=0}^{\infty} nA_n W_n(x) + \sum_{n=0}^{\infty} n^2 A_n W_n(x).$$

The uniform convergence of the series involved in this last relation shows that the differentiation was legitimate and establishes the existence and continuity of $S^{(4)}(x)$. In the same manner successive differentiations and substitutions establish the existence of the derivatives of even order up to $S^{(2k)}(x)$. It is only necessary to eliminate by means of preceding equations all of the series which appear with coefficients involving x, since otherwise the double differentiation would introduce terms containing $W_n'(x)$, which cannot be expressed in terms of $W_n(x)$.

COROLLARY. *In a fixed finite interval Theorem* IV *may be applied to the series* (3) *without change of wording.*

Since we may take our interval as large as we like, it appears that the corollary establishes the existence of derivatives up to order $2k$ for all values of x provided the hypotheses hold uniformly in the infinite interval.

Our second theorem furnishing necessary conditions deals with the behavior of the function at infinity, and is as follows:

THEOREM V. *If the series*

$$S(x) = A_0W_0(x) + A_1W_1(x) + A_2W_2(x) + A_3W_3(x) + \cdots$$

converges so that the remainder after n terms is $O(n^{-k/2})$ *uniformly in the infinite interval, then*

$$S(x) = O(|x|^{-k})$$

when $|x|$ *is large.*

It may be verified without difficulty that the functions $W_n(x)$, $W'_{n-1}(x)$, $W_{n-2}(x)$ all have the same sign when $x>h$, so that from the relationship

$$n^{1/2}W_n(x) - (n-1)^{1/2}W_{n-2}(x) = 2W'_{n-1}(x),$$

obtained from (11), we derive the inequality

$$|W_{n-2}(x)| < n^{1/2}(n-1)^{-1/2}|W_n(x)|.$$

By successive repetitions this gives us

$$|W_{n-2m}(x)| < [n(n-2)\cdots(n-2m+2)]^{1/2} \cdot [(n-1)(n-3)\cdots(n-2m+1)]^{-1/2}|W_n(x)| < n^{1/2}|W_n(x)|.$$

Likewise

$$|W_{n-2m+1}(x)| < n^{1/2}|W_{n-1}(x)|.$$

Now if we let $X=h+2k+6$ and choose $|x|\geqq X$ we may apply inequality (29), with the exponent $-k/2-3/2$, to the functions $W_{n-1}(x)$ and $W_n(x)$, so that in view of the inequalities just obtained above

$$W_m(x) = O(n^{-k/2-1}),\quad m = 0,1,2,\cdots,n,\quad |x|\geqq X, \tag{42}$$

and from this follows

$$S_n(x) = O(n^{-k/2}).$$

By hypothesis

$$S(x) = S_n(x) + O(n^{-k/2}),$$

so that

$$S(x) = O(n^{-k/2}) = O(X^{-k}).$$

Since this result is true in particular when $|x|=X$, we easily draw the conclusion of the theorem. The fact that X takes on isolated values only does not affect the truth of the theorem for the continuous variable x.

9. Up to the present we have been concerned with the degree of convergence of the expansions (3) or (39) in which the coefficients are determined by the formulas (4) or (40). Now, however, we take up the case in which the function $f(x)$ is to be expressed as a linear combination of $W_0, W_1, \cdots, W_n$, with coefficients which are not necessarily given by (40). The first result is

THEOREM VI. *If $f(x)$ satisfies a Lipschitz condition*

$$|f(x_1) - f(x_2)| < L|x_1 - x_2|$$

and if

$$|xf(x)| < L,$$

then there exists a sum

$$\Sigma_n(x) = A_{0n}W_0(x) + A_{1n}W_1(x) + \cdots + A_{nn}W_n(x) \tag{43}$$

such that

$$|f(x) - \Sigma_n(x)| < BLn^{-1/2},$$

in which B denotes a constant independent of n and of f(x).

We shall prove the theorem only for even values of n, $n=2m$, though it will be clear that the result holds without this restriction. Our starting point is a theorem due to Jackson* which asserts the existence of a polynomial $P_{1m}(x)$ of degree m at most such that in a closed interval of length J

$$|f(x) - P_{1m}(x)| < B_1 LJm^{-1}.$$

Here as in subsequent formulas B_1, B_2, etc. denote constants independent of n and of $f(x)$. We shall choose the interval

$$-X \leqq x \leqq X, \quad X = h + 8,$$

so that

$$J = 2[(8m+2)^{1/2} + 8] = (32m)^{1/2}[1 + O(m^{-1/2})].$$

We may therefore write

$$P_{1m}(x) = f(x) + \rho_{1m}(x), \tag{44}$$

where

$$|\rho_{1m}(x)| < B_2 Lm^{-1/2}.$$

We next turn to the expansion of unity in a series $S_m(x)$ as given by (41). Clearly we have

$$S_m(x) = e^{-x^2/4} P_{2m}(x),$$

where $P_{2m}(x)$ is a polynomial of degree m at most, so that (41) can be put in the form

$$e^{-x^2/4} P_{2m}(x) = 1 + \rho_{2m}(x) \tag{45}$$

with the conditions

$$|\rho_{2m}(x)| < B_3 m^{-1/2}, \qquad -N \leqq x \leqq N,$$

and

$$|\rho_{2m}(x)| < B_4, \qquad -\infty < x < \infty.$$

Now we multiply (44) and (45), obtaining

$$e^{-x^2/4} P_n(x) = f(x) + R_n(x),$$

in which $P_n(x)$ is a polynomial of degree n at most. From the inequalities satisfied by $\rho_{1n}(x)$ and $\rho_{2n}(x)$ together with the fact that $|xf(x)| < L$ we readily deduce that

* These Transactions, vol. 14 (1913), p. 351, Theorem V.

$$|R_n(x)| < BLn^{-1/2}. \tag{46}$$

Since $e^{-x^2/4}P_n(x)$ is a linear combination of $W_0, W_1, \cdots, W_n$, we may set

$$e^{-x^2/4}P_n(x) = \Sigma_n(x)$$

and have

$$\Sigma_n(x) = f(x) + R_n(x) \tag{47}$$

in the interval $-X \leqq x \leqq X$, so that the desired theorem is established in this interval.

Now we consider the magnitude of the remainder term for x outside of this interval. First of all we observe by means of (42) for the case $k=0$, that

$$\int_{-X}^{X} W_i(x)W_j(x)dx = \begin{cases} 1 + O(n^{-2}), & i = j \leqq n, \\ O(n^{-2}), & i < j \leqq n. \end{cases}$$

If A_{jn} denotes the largest coefficient in (43) we multiply (47) by $W_j(x)$, integrate from $-X$ to X, and have

$$A_{jn}(1 + O(n^{-1})) = \int_{-X}^{X} f(x)W_j(x)dx + \int_{-X}^{X} W_j(x)R_n(x)dx,$$

from which it appears that A_{jn} is bounded with respect to n. For from (26) and (29) it is seen that $|x^{1/6}W_j(x)|$ is bounded, and the same is true of $|x\,f(x)|$, so that the first integral is bounded. That the second integral is bounded is obvious from (46).

Finally we use (42) again for the case $k=1$ getting

$$|W_p(x)| < B_5 n^{-3/2} \qquad (p = 0, 1, \cdots, n),$$

whence

$$|\Sigma_n(x)| < B_6 L n^{-1/2},$$

and since $|f(x)| < Ln^{-1/2}$ when $|x| > X$, we see that, for a suitable B, (46) holds for all values of x, and the theorem is established for the infinite interval.

If x is replaced by $2^{1/2}x$ a sum of the type (43) is converted into a sum of the type

$$\sigma_n(x) = b_0\phi_0(x) + b_1\phi_1(x) + \cdots + b_n\phi_n(x), \tag{48}$$

whence we have

THEOREM VII. *Theorem* VI *also holds for the sum* (48).

On the other hand the introduction of the factor $e^{-x^2/4}$ enables us to deduce from Theorem VI the following

THEOREM VIII. *If $\phi(x)$ satisfies a condition of the form*

$$| e^{x_1^2/4}\phi(x_1) - e^{x_2^2/4}\phi(x_2) | < L | x_1 - x_2 |$$

and if

$$| xe^{x^2/4}\phi(x) | < L$$

there exists a sum of the type (48) *such that*

$$| \phi(x) - \sigma_n(x) | < BLn^{-1/2}e^{-x^2/4}.$$

10. We are now in a position to establish a theorem very similar to one proved by Jackson* for the indefinite integral of a trigonometric sum.

THEOREM IX. *Let $\phi(x)$ be an integrable function for which*

$$\int_{-\infty}^{\infty} \phi(x)dx = 0, \tag{49}$$

and suppose that there exists a sum $\sigma_n(x)$ of the form given by (48) *such that*

$$\phi(x) = \sigma_n(x) + \rho_n(x), \tag{50}$$

where

$$| \rho_n(x) | < \epsilon e^{-x^2/4}$$

for all values of x in the infinite interval. Then

$$\int_{-\infty}^{x} \phi(x)dx = c_0\phi_0(x) + c_1\phi_1(x) + \cdots + c_n\phi_n(x) + r_n(x) \tag{51}$$

in which

$$| r_n(x) | < 5B\epsilon n^{-1/2}e^{-x^2/4}.$$

The proof is so similar to that of Jackson that it might be omitted, were it not that the behavior of the remainder terms when x is large requires some slight additional investigation. We shall sketch the proof briefly. If we integrate (50) from $-\infty$ to ∞ and use (49) we find that the term $b_0\phi_0$ can be included with $\rho_n(x)$ to form a new remainder $\rho_n'(x)$ for which

$$| \rho_n'(x) | < (2^{1/2} + 1)\epsilon e^{-x^2/4}.$$

We then integrate (50) from $-\infty$ to x, obtaining

$$\int_{-\infty}^{x} \phi(x)dx = b_1\phi_0(x) + b_2\phi_1(x) + \cdots + b_n\phi_{n-1}(x) + R_n(x) \tag{52}$$

in which

$$R_n(x) = \int_{-\infty}^{x} \rho_n'(x)dx.$$

* These Transactions, vol. 13 (1912), pp. 491–515.

Clearly $R_n(x)$ vanishes at $-\infty$ and in view of (49) it also vanishes at ∞. Therefore

$$R_n(x) = -\int_x^\infty \rho_n'(x)dx.$$

If x is negative we use the first form of $R_n(x)$, but if x is positive we use the second form. Suppose that x is positive. Then

$$\begin{aligned} |R_n(x)| &< (2^{1/2}+1)\epsilon \int_x^\infty e^{-x^2/4}dx \\ &< (2^{1/2}+1)\epsilon x^{-1}\int_x^\infty e^{-x^2/4}xdx \\ &= 2(2^{1/2}+1)\epsilon x^{-1}e^{-x^2/4} < 5\epsilon x^{-1}e^{-x^2/4}. \end{aligned}$$

The same is true when x is negative. Thus we have shown that $R_n(x)$ satisfies the second part of the hypothesis of Theorem VIII with $L=5\epsilon$. That $R_n(x)$ satisfies the first part of the hypothesis with $L=5\epsilon$ follows from the fact, readily shown from the foregoing inequalities, that the derivative of $e^{x^2/4}R_n(x)$ does not exceed 5ϵ in absolute value.

Therefore from Theorem VIII we see that there exists a sum $s_n(x)$ involving the first n ϕ's, such that

$$|R_n(x) - s_n(x)| < 5B\epsilon n^{-1/2}e^{-x^2/4}.$$

We use this result in connection with (52), combine the two sums and obtain (51).

From the two theorems VIII and IX we derive

THEOREM X. *If $\phi(x)$ has a $(k-1)$th derivative which satisfies the conditions*

$$|e^{x_1^2/4}\phi^{(k-1)}(x_1) - e^{x_2^2/4}\phi^{(k-1)}(x_2)| < L|x_1 - x_2|,$$

$$|xe^{x^2/4}\phi^{(k-1)}(x)| < L, \tag{53}$$

and if

$$\lim_{x=\pm\infty} \phi(x) = 0, \tag{54}$$

then there exists a sum $\sigma_n(x)$ of the form (48) *such that*

$$|\phi(x) - \sigma_n(x)| < (5B)^k L n^{-k/2}e^{-x^2/4}.$$

To establish this we first apply Theorem VIII to the function $\phi^{(k-1)}(x)$, then use Theorem IX to obtain an approximating sum for $\phi^{(k-2)}(x)$, again use IX to obtain an approximating sum for $\phi^{(k-3)}(x)$, and continue thus till

the desired result is reached for $\phi(x)$. In making these applications of Theorem IX we need to know, in addition to the stated hypotheses, that

$$\int_{-\infty}^{\infty} \phi^{(m)}(x)dx = 0 \qquad (m = 1,2, \cdots, k-1),$$

or what amounts to the same thing, that

$$\lim_{x=\pm\infty} \phi^{(m)}(x) = 0 \qquad (m = 0,1,2, \cdots, k-2).$$

Now if x is large and positive

$$\phi^{(k-2)} = C - \int_x^{\infty} \phi^{(k-1)}(x)dx$$

$$= C + O(x^{-2}e^{-x^2/2}),$$

$$\phi^{(k-3)} = Cx + C' + O(x^{-3}e^{-x^2/2}).$$

Continuing thus we find that $\phi(x)$ may be expressed as a polynomial with an added term which is $O(x^{-k}e^{-x^2/2})$. But the condition (54) shows that the polynomial must vanish identically, and therefore C, C', etc. all vanish, which completes the proof.

If in Theorem X we make the change of function

$$\phi(x) = e^{-x^2/4}f(x),$$

we are enabled to state a theorem regarding approximation by sums of the type (41):

THEOREM XI. *If $f(x)$ has a $(k-1)$th derivative which satisfies a Lipschitz condition*

$$|f^{(k-1)}(x_1) - f^{(k-1)}(x_2)| < L|x_1 - x_2|,$$

and if

$$|x^{k-m}f^{(m)}(x)| < L \qquad (m = 0,1,2, \cdots, k-1),$$

then there exists a sum $\Sigma_n(x)$ such that

$$\Sigma_n(x) = f(x) + O(n^{-k/2})$$

uniformly in the infinite interval.

The proof of this theorem may be left to the reader.

STANFORD UNIVERSITY,
STANFORD UNIVERSITY, CALIF.

FOUCAULT'S PENDULUM IN ELLIPTIC SPACE*

BY
JAMES PIERPONT

1. In the following for e- read euclidean, for E- read elliptic. Let x, y, z be ordinary rectangular coördinates of a point in e-space whose origin is O. Set $r^2=x^2+y^2+z^2$, $\lambda=4R^2+r^2$, $\mu=4R^2-r^2$, where R is an arbitrary positive constant. For all points of e-space

$$d\sigma^2 = dx^2 + dy^2 + dz^2.$$

For points within and on the e-sphere $\mu=0$ we establish an elliptic metric by means of

$$ds = (4R^2/\lambda)d\sigma. \tag{1}$$

Points outside of $\mu=0$ do not exist in E-space while two diametral points on $\mu=0$ are regarded as identical.

An E-straight is an e-circle cutting $\mu=0$ in diametral points; an E-plane is an e-sphere cutting $\mu=0$ along a great circle. The e-sphere $\mu=0$ is regarded as an E-plane. Angles between E-straights and planes have the same measure in E- as in e-space.

The 4 E-planes $x=0$, $y=0$, $z=0$, $\mu=0$ form an E-tetrahedron which we call τ. From a point xyz drop E-perpendiculars on the 4 faces of τ and let δ_i, $i=1, 2, 3, 4$, be their E-lengths. We set

$$z_i = R \sin(\delta_i/R).$$

We find

$$z_1 = 4R^2x/\lambda, \quad z_2 = 4R^2y/\lambda, \quad z_3 = 4R^2z/\lambda, \quad z_4 = R\mu/\lambda.$$

Also

$$z_1^2 + z_2^2 + z_3^2 + z_4^2 = R^2, \qquad ds^2 = dz_1^2 + dz_2^2 + dz_3^2 + dz_4^2. \tag{2}$$

In these coördinates the equation of an E-plane has the form

$$a_1z_1 + a_2z_2 + a_3z_3 + a_4z_4 = 0.$$

The distance δ between two points z, z' is given by

$$\cos(\delta/R) = \frac{z_1z_1' + z_2z_2' + z_3z_3' + z_4z_4'}{R^2}.$$

---

* Presented to the Society, February 23, 1929; received by the editors February 1, 1929.

We may without loss of generality set $R=1$ and this will be done in the following.

2. Let $c_1, \cdots, c_4$ be the coördinates of the point of suspension O' whose latitude is ϕ and whose longitude is θ. Let $\overline{OO'}=\rho$ in E-measure. For brevity we set

$$r = \sin\rho, \quad r' = \cos\rho, \quad r_1 = \cot\rho\,; \quad p = \sin\phi, \quad p' = \cos\phi.$$

Then

$$c_1 = rp'\cos\theta, \quad c_2 = rp'\sin\theta, \quad c_3 = rp, \quad c_4 = r'.$$

Let us now displace the xyz axes so that O moves to O'. The new e-axes call ξ, η, ζ, where $+\xi$, $+\eta$ point south and east respectively, while $+\zeta$ points to the zenith. These axes define a new E-tetrahedron which we call τ'.

The relation between the coördinates $z_1, \cdots, z_4$ referred to τ and the coördinates $\zeta_1, \cdots, \zeta_4$ of the same point referred to τ' is given by the table, read as in ordinary analytic geometry.

(3)

	z_1	z_2	z_3	z_4
ζ_1	$p\cos\theta$	$p\sin\theta$	$-p'$	0
ζ_2	$-\sin\theta$	$\cos\theta$	0	0
ζ_3	$r'p'\cos\theta$	$r'p'\sin\theta$	$r'p$	$-r$
ζ_4	$rp'\cos\theta$	$rp'\sin\theta$	rp	r'

We now suppose that τ remains fixed in space, that the earth rotates about the z axis with a constant angular velocity $k=\dot\theta=d\theta/dt$ and that finally τ' is rigidly attached to the earth.

We suppose the bob B of the pendulum to be a particle of mass m, and attached to the point of suspension c or O' by a weightless rod of length L in E-measure. Set $l=\sin L$, $l'=\cos L$; let the plane through B and the ζ axis make the angle ω with the $\xi\cdot\zeta$ plane, let the rod $O'B$ make with the negative ζ axis the angle ψ. Then the coördinates of B relative to τ' are

$$\zeta_1 = l\sin\psi\cos\omega, \quad \zeta_2 = l\sin\psi\sin\omega, \quad \zeta_3 = -l\cos\psi, \quad \zeta_4 = l'. \tag{4}$$

3. Let the force F act on a particle; if the particle is displaced along an elementary segment of length ds as defined by (1) or by (2) and if θ is the angle between F and ds we assume with Killing* that the work done is $dW=F\cos\theta ds$. We ask now what is dW when ψ receives the increment $d\psi$. In the triangle $OO'B$ we have setting $\overline{OB}=\beta$ in E-measure

* W. Killing, *Die Mechanik in den nicht-euklidischen Raumformen*, Crelle's Journal, vol. 98 (1885), p. 1.

$$\sin B = \frac{\sin \rho}{\sin \beta}; \quad \sin \psi = -\cos \theta.$$

As $ds = \sin L d\psi$ we have

$$dW = -F \frac{\sin \rho}{\sin \beta} \sin L \sin \psi d\psi = -F \frac{\sin \rho}{\sin \beta} d\zeta_3. \tag{5}$$

Since the length of the pendulum L is negligible compared with ρ, $\sin \beta = \sin \rho$ with a high degree of exactitude. We may therefore write

$$dW = -F d\zeta_3 = -F \sin L \sin \psi d\psi, \tag{6}$$

which is what we would expect at once.

We note that the work done when ω receives an increment is 0, since in this case $\theta = \pi/2$, hence $\partial W/\partial \omega = 0$.

4. We now wish to calculate the velocity v of the bob B. We have

$$v^2 = \dot{s}^2 = \dot{z}_1^2 + \dot{z}_2^2 + \dot{z}_3^2 + \dot{z}_4^2.$$

From the table (3) we express the z's in terms of the ζ's and these by means of (4) in terms of ψ, ω. We then differentiate the z's, squared, and add. We find, setting as before $k = \dot{\theta}$,

$$\begin{aligned} v^2 = {} & k^2[l^2 \sin^2 \psi \sin^2 \omega + (pl \sin \psi \cos \omega - r'p'l \cos \psi + l'p'r)^2] \\ & + l^2\dot{\psi}^2 + l^2 \sin^2 \psi \dot{\omega}^2 \\ & + 2k\dot{\psi}[rp'll' \cos \psi \sin \omega - l^2 r'p' \sin \omega] \\ & + 2k\dot{\omega}[l^2 p \sin^2 \psi - r'p'l^2 \sin \psi \cos \psi \cos \omega + rp'll' \sin \psi \cos \omega]. \end{aligned} \tag{7}$$

The kinetic energy of the bob B we define by

$$T = \tfrac{1}{2} m v^2.$$

5. We assume now that the motion of the bob B takes place according to Hamilton's principle

$$\int (\delta T + \delta W) dt = 0.$$

On performing the variation we get as usual Lagrange's equation

$$\frac{d}{dt} \frac{\partial T}{\partial \dot{\omega}} - \frac{\partial T}{\partial \omega} = 0, \qquad \frac{d}{dt} \frac{\partial T}{\partial \dot{\psi}} - \frac{\partial T}{\partial \psi} = \frac{\partial W}{\partial \psi}. \tag{8}$$

Let us calculate the ω equation. From (7)

$$\frac{\partial T}{\partial \dot{\omega}} = l^2(\sin^2 \psi) \cdot \dot{\omega} + k[l^2 p \sin^2 \psi - r'p'l^2 \sin \psi \cos \psi \cos \omega + rp'll' \sin \psi \cos \omega],$$

$$\frac{\partial T}{\partial \omega} = k^2[l^2 \sin^2\psi \sin\omega\cos\omega - pl\sin\psi\sin\omega(pl\sin\psi\cos\omega - r'p'l\cos\psi + l'p'r)]$$
$$+ k\psi[rp'll'\cos\psi\cos\omega - l^2r'p'\cos\omega]$$
$$+ k\dot{\omega}[r'p'l^2\sin\psi\cos\psi\sin\omega - rp'll'\sin\psi\sin\omega].$$

Thus the first equation (8) gives

$$l^2\frac{d}{dt}(\dot{\omega}\sin^2\psi) + kl^2\frac{d}{dt}(\sin^2\psi) - kr'p'\frac{l^2}{2}\frac{d}{dt}(\sin 2\psi\cos\omega) + krp'll'\frac{d}{dt}(\sin\psi\cos\omega)$$
$$= k^2l^2\sin^2\psi\sin\omega\cos\omega - k^2pl\sin\psi\sin\omega(pl\sin\psi\cos\omega - r'p'l\cos\psi + l'p'r)$$
$$+ krp'll'((\cos\omega\cos\psi)\dot{\psi} - (\sin\psi\sin\omega)\dot{\omega})$$
$$- kl^2r'p'((\cos\omega)\dot{\psi} - (\sin\psi\cos\psi\sin\omega)\dot{\omega}). \tag{9}$$

We will now suppose that ψ is so small that we may set $\sin\psi = \psi$ without sensible error; then (9) becomes

$$l^2\frac{d}{dt}(\dot{\omega}\psi^2) + kl^2p\frac{d}{dt}(\psi^2) - kr'p'l^2\frac{d}{dt}(\psi\cos\omega) + krp'll'\frac{d}{dt}(\psi\cos\omega)$$
$$= k^2l^2\psi^2\sin\omega\cos\omega - k^2pl\psi\sin\omega\{pl\psi\cos\omega + (l'r - r'l)p'\}$$
$$+ krp'll'\frac{d}{dt}(\psi\cos\omega) - kl^2r'p'\frac{d}{dt}(\psi\cos\omega);$$

or as $l'r - r'l = \cos L\sin\rho - \cos\rho\sin L = \sin(\rho - L) = \sin\rho = r$ very nearly, we get

$$l(\ddot{\omega}\psi^2 + 2\psi\dot{\omega}\dot{\psi} + 2kp\psi\dot{\psi})$$
$$= k^2l\psi^2\sin\omega\cos\omega - k^2p^2l\psi^2\sin\omega\cos\omega - k^2pp'r\psi\sin\omega.$$

Hence

$$2\dot{\psi}(\dot{\omega} + kp) + \psi\ddot{\omega} = k^2p'^2\psi\sin\omega\cos\omega - (k^2pp'r/l)\sin\omega.$$

These are entirely analogous to the equations of classical mechanics. Under similar conditions we may say therefore that in first approximation the angular velocity of the plane of vibration is

$$\dot{\omega} = -k\sin\phi.$$

YALE UNIVERSITY,
NEW HAVEN, CONN.

ON A CERTAIN FORMULA OF MECHANICAL QUADRATURES WITH NON-EQUIDISTANT ORDINATES*

BY

J. A. SHOHAT (JACQUES CHOKHATE)

Introduction. The object of this paper is to investigate the formula of mechanical quadratures

$$\text{(A)} \qquad \int_{-n}^{n} u_x dx = a(u_{-n} + u_n) + \sum_{i=1}^{s} a_i(u_{-n_i} + u_{n_i}) \qquad (u_x \equiv u(x)),$$

$$\left[\text{(A')} \quad \int_{0}^{2n} u_x dx = a(u_0 + u_{2n}) + \sum_{i=1}^{s} a_i(u_{n-n_i} + u_{n+n_i})\right],$$

$$\text{(1)} \qquad 0 < n_1 < n_2 < \cdots < n_s < n,$$

where n is a certain given *finite* quantity, $u(x)$ is a bounded (R)-integrable function defined on $(-n, n)$, a, a_i, $n_i (i = 1, 2, \cdots, s)$ are to be determined, making use of certain conditions to be stated later.

This formula has been suggested by G. F. Hardy† and given, for $s = 2$, in King's *Text-Book*.‡ A discussion of it, for $s = 2, 3$, has been given by J. W. Glover.§ Since the formula under consideration is being frequently used and yields a very close approximation, it seems to be of interest to investigate its theoretical basis for any positive integral value of s. The following questions naturally arise.

(i) *Existence of formula* (A) *for any* $s = 1, 2, 3, \cdots$; i.e., can we always find in (A) *real* a, a_i, $n_i (i = 1, 2, \cdots, s)$ satisfying (1)?

(ii) *Theoretical basis of* (A), i.e., find an algorithm yielding the values of the constants enumerated in (i), for any positive integer s.

(iii) *Convergence of formula* (A), i.e. investigate

$$\lim_{s \to \infty} \left[\int_{-n}^{n} u_x dx - a(u_{-n} + u_n) - \sum_{i=1}^{s} a_i(u_{-n_i} + u_{n_i}) \right].$$

(iv) *Remainder and accuracy of formula* (A).

---

* Presented to the Society, March 29, 1929; received by the editors January 31, 1929.

† G. F. Hardy, *On some formulas of approximate summation*, Journal of the Institute of Actuaries and Assurance Magazine (London), vol. 24 (1884), pp. 95–110; p. 95.

‡ G. King, *Institute of Actuaries Text-Book*, 2d edition, London, 1902, part II, pp. 489–490.

§ J. W. Glover, *Quadrature formulae when ordinates are not equidistant*, Proceedings of the International Mathematical Congress, Toronto, 1924, vol. II, pp. 831–835.

In what follows we shall endeavor to answer the aforesaid questions, and to show that *formula* (A) *is closely related to the theory of Jacobi polynomials.*

1. **Statement of the problem.** *Find* $2s+1$ *real constants*: $a, a_1, a_2, \cdots, a_s, (0<)n_1<n_2< \cdots <n_s(<n)$, *such that formula* (A) *be exact for an arbitrary polynomial* $G_{4s+1}(s)$ *of degree* $\leqq 4s+1$. [Hereafter, $G_m(x)\equiv\sum_{i=0}^{m}g_ix^i$ generally stands for an *arbitrary* polynomial of degree $\leqq m$.]

Formula (A) obviously holds true for $x^{2k+1}(k=0, 1, 2, \cdots)$. Hence, it is necessary and sufficient that

$$\int_{-n}^{n} x^{2k}dx = \frac{2n^{2k+1}}{2k+1} = 2an^{2k} + 2\sum_{i=1}^{s} a_i n_i^{2k} \qquad (k = 0,1, \cdots, 2s),$$

which leads to the following system of $2s+1$ equations:

$$\sum_{j=0}^{s} b_j \alpha_j^k = \frac{1}{2k+1} \qquad (k = 0,1, \cdots, 2s), \tag{2}$$

with

$$\alpha_j = (n_j/n)^2, \quad b_j = a_j/n \quad (j = 0,1, \cdots, s\,;\, n_0 = n, \alpha_0 = 1,\ a_0 = a). \tag{3}$$

2. **Application of continued fractions. Relation to Jacobi polynomials.** Introduce the series (finite or infinite, if s be allowed to increase indefinitely)

$$K(x) \equiv \sum_{m=0} \frac{s_m}{x^{m+1}} \qquad \left(s_m = \frac{1}{2m+1} - \frac{1}{2m+3} = \frac{2}{(2m+1)(2m+3)}\right), \tag{4}$$

$$s_m = \int_{-1}^{1} (1-x^2)x^{2m}dx = \int_0^1 \frac{x^{-1/2}(1-x)}{2} x^m dx \qquad (m = 0,1, \cdots). \tag{5}$$

(5) yields the following integral representation of $K(x)$:

$$K(x) = \int_0^1 \frac{\frac{1}{2}y^{-1/2}(1-y)}{x-y}dy. \tag{6}$$

In the system (2) multiply the kth equation by $g_k(k=0, 1, 2, \cdots, 2s)$ and add. Then (2) becomes equivalent to

$$\omega(G_{2s}) = \sum_{j=0}^{s} b_j G_{2s}(\alpha_j), \tag{7}$$

where, in general,

$$\omega(G_m) = \sum_{k=0}^{m} g_k/(2k+1). \tag{8}$$

On the other hand, making use of

$$(9)\qquad p_1(x) = \tfrac{1}{2}x^{-1/2}, \quad \int_0^1 p_1(x)x^m dx = 1/(2m+1) \qquad (m = 0,1,\cdots),$$

we can write (8) in integral form as follows:

$$(10)\qquad \omega(G_{2s}) = \int_0^1 p_1(x)G_{2s}(x)dx.$$

Apply (7), (10) to the polynomial

$$(11)\qquad \Psi_j(x) = \frac{\Psi(x)}{x-\alpha_j} \qquad (j = 0,1,\cdots,s),$$

$$\Psi(x) = \prod_{j=0}^{s}(x-\alpha_j) \equiv (x-1)\Phi_s(x), \quad \Phi_s(x) = \prod_{i=1}^{s}(x-\alpha_i):$$

$$(12)\qquad b_j = \int_0^1 \frac{p_1(x)\Psi(x)dx}{(x-\alpha_j)\Psi'(\alpha_j)},$$

$$(13)\qquad b_j = \frac{R(\alpha_j)}{\Psi'(\alpha_j)}, \qquad R(x) = \int_0^1 p_1(y)\frac{\Psi(x)-\Psi(y)}{x-y}dy \quad (j = 0,1,\cdots,s);$$

$$(14)\qquad b_0 = \int_0^1 p_1(x)\frac{\Phi_s(x)dx}{\Phi_s(1)}, \quad b_i = \int_0^1 \frac{p_1(x)(1-x)\Phi_s(x)dx}{(1-\alpha_i)(x-\alpha_i)\Phi_s'(\alpha_i)} \quad (i=1,2,\cdots,s).$$

The same formulas (7), (10), applied to $\Psi(x)G_{s-1}(x)$, give

$$(15)\qquad \omega[\Psi(x)G_{s-1}(x)] = \int_0^1 p_1(x)\Psi(x)G_{s-1}(x)dx = 0,$$

$$(16)\qquad \int_0^1 p(x)\Phi_s(x)G_{s-1}(x)dx = 0, \quad p(x) = \frac{1}{2}x^{-1/2}(1-x).$$

The relations (6), (7), (16) *are fundamental.* (16), equivalent to $\int_0^1 p(x)\Phi_s(x)\cdot x^k dx = 0$ $(k=0,1,\cdots,s)$, shows that *in the product*

$$\Phi_s(x)K(x) \equiv \Phi_s(x)\int_0^1 \frac{p(y)dy}{x-y}$$

the terms in $1/x, 1/x^2, \cdots, 1/x^s$ *are absent,* i.e.

$$(17)\qquad \Phi_s(x)\int_0^1 \frac{p(y)dy}{x-y} = P_s(x)(\text{polynomial of degree } s-1) + \frac{1}{(x^{s+1})}.^*$$

* We write, in general,

$$\frac{d_1}{x^p} + \frac{d_2}{x^{p+1}} + \cdots \equiv \left(\frac{1}{x^p}\right) \qquad (p > 0).$$

Hence,* $P_s(x)/\Phi_s(x)$ *is a convergent to the continued fraction*

$$\int_0^1 \frac{p(y)dy}{x-y} = \frac{\lambda_1|}{|x-c_1} - \frac{\lambda_2|}{|x-c_2} - \cdots \tag{18}$$

which necessarily exists, since $p(x)$ *is not negative in* $(0, 1)$.† Furthermore, (18) *is a special case of the continued fraction "associated"* † *with the integral*

$$\int_0^1 \frac{y^{\alpha-1}(1-y)^{\beta-1}dy}{x-y} \qquad (\alpha,\beta > 0),$$

for $\alpha=\frac{1}{2}$, $\beta=2$, *where the denominators of the convergents are Jacobi polynomials*, which, in turn, are but a particular case of orthogonal Tchebycheff polynomials.

3. **Some general properties of orthogonal Tchebycheff polynomials. Polynomials of Jacobi.** Any "*c-function*" $p(x)$ defined on the finite or infinite interval (a, b), i.e. *non-negative and having all the "moments"* $\int_a^b p(x)x^n dx$ $(n=0, 1, \cdots)$, gives rise, it is known, to a system of orthogonal Tchebycheff polynomials

$$\Phi_n[p(x)\ ;\ a,b,x] \equiv \Phi_n(x) = x^n - S_n x^{n-1} + \cdots \qquad (n = 0,1,2, \cdots) \tag{19}$$

uniquely determined by either one of the equivalent set of relations

$$\begin{aligned} &\int_a^b p(x)\Phi_n(x)\Phi_m(x)dx = 0 && (m \neq n\ ;\ m,n = 0,1,2, \cdots), \\ &\int_a^b p(x)\Phi_n(x)G_{n-1}(x)dx = 0 && (n = 0,1,2, \cdots). \end{aligned} \tag{20}$$

We can *normalize* the system (19), and we get the sequence of polynomials

$$\phi_n[p(x)\ ;\ a,b\ ;\ x] \equiv \phi_n(p\ ;\ x) \equiv \phi_n(x) = a_n(p)x^n + \cdots \qquad (a_n(p) \equiv a_n > 0\ ;\ \ n = 0,1, \cdots), \tag{21}$$

$$\int_a^b p(x)\phi_n(x)\phi_m(x)dx = \begin{cases} 0 & (n \neq m), \\ 1 & (n = m). \end{cases} \tag{22}$$

The following are some of the most important properties of $\{\Phi_n(x)\}$.‡

---

* Cf., for example, Jordan, *Cours d'Analyse*, 2d edition (1893), vol. I, p. 373.

† O. Perron, *Die Lehre von den Kettenbrüchen*, Leipzig and Berlin, Teubner, 1913, p. 377.

‡ Cf., for example, J. Chokhate, *Sur le développement de l'intégrale* $\int_a^b [p(x)/(x-y)]dy$ *en fraction continue et sur les polynomes de Tchebycheff*, Rendiconti del Circolo Matematico di Palermo, vol. 47 (1923), pp. 25–46; pp. 25–30.

(i) $\{\Phi_n(x)\}$ are the denominators of the convergents $\{\Psi_n(x)/\Phi_n(x)\}$ to the "associated" continued fraction

$$\int_a^b \frac{p(y)dy}{x-y} = \frac{\lambda_1|}{|x-c_1} - \frac{\lambda_2|}{|x-c_2} - \cdots \qquad (\lambda_i, c_i = \text{const.};\ \lambda_i > 0).$$

(ii) $$\Phi_{n+1}(x) = (x - c_{n+1})\Phi_n(x) - \lambda_{n+1}\Phi_{n-1}(x) \qquad (n \geqq 1),$$

$$\Psi_{n+1}(x)\Phi_n(x) - \Psi_n(x)\Phi_{n+1}(x) = \lambda_1\lambda_2 \cdots \lambda_{n+1} = \frac{1}{a_n^2}.$$

(iii)
$$\lambda_n = \frac{a_{n-2}^2}{a_{n-1}^2} \qquad \left(n \geqq 2\ ;\ \lambda_1 = \int_a^b p(x)dx\right);$$

$$c_n = S_n - S_{n-1} \qquad \left(n \geqq 2\ ;\ c_1 = \frac{\int_a^b xp(x)dx}{\int_a^b p(x)dx}\right).$$

(iv) $$\sum_{i=0}^{n}\phi_i^2(x) \equiv \sum_{i=0}^{n}a_i^2\Phi_i(x) = \frac{a_n}{a_{n+1}}[\phi_{n+1}'(x)\phi_n(x) - \phi_{n+1}(x)\phi_n'(x)].$$

(v) $\Phi_n(x)$ has all roots $x_{i,n}(i=1, 2, \cdots, n)$ real, distinct and between (a, b).

(vi) $$x_{1,n+1} < x_{1,n} < x_{2,n+1} < x_{2,n} < \cdots < x_{n,n} < x_{n+1,n+1}.$$

(vii) For $n \to \infty$: $\lim x_{1,n} = a$, $\lim x_{n,n} = b$,

assuming the non-existence of numbers α, β such that

$$\int_a^\alpha p(x)dx = 0, \quad \int_\beta^b p(x)dx = 0 \qquad (a \leqq \alpha, \beta \leqq b).$$

(viii) In case of (a, b) finite* we find, for n sufficiently large, roots $x_{i,n}$ in any sub-interval (c, d) such that $\int_c^d p(x)dx > 0$ $(a \leqq c < d \leqq b)$.

(ix) If $a = -b$, and $p(x) \equiv p(-x)$, then ("*symmetric*" *Tchebycheff polynomials*):

(23)
$$\phi_{2n+\epsilon}[p(x)\ ;\ -b, b\ ;\ x] \equiv x^\epsilon\phi_n[p(x^{1/2})x^{\epsilon-1/2}\ ;\ 0, b^2\ ;\ x^2] \qquad (\epsilon = 0, 1);$$
$$c_n = 0\ ; \qquad x_{i,n} + x_{n-i+1} = 0 \qquad (i = 1, 2, \cdots, n\ ;\ n = 1, 2, \cdots).$$

If we take, in the formulas above,

---

* The case of (a, b) infinite requires the consideration of the nature of $p(x)$ for $|x|$ very large.

(24) (a, b) finite, say $(0, 1)$, $p(x) = x^{\alpha-1}(1-x)^{\beta-1}$ $(\alpha, \beta > 0)$,

we get *polynomials of Jacobi* (Legendre polynomials: $\alpha = \beta = 1$; trigonometric polynomials: $\alpha = \beta = \frac{1}{2}$) which we denote by

$$\Phi_n(\alpha,\beta;0,1\,;\,x) \equiv \Phi_n(\alpha,\beta\,;\,x) \equiv \Phi_n(x) = x^n - S_n x^{n-1} + \cdots \qquad (n = 0,1,2,\cdots). \tag{25}$$

Here are some of their properties to be used later.*

$$\Phi_n(\alpha,\beta;0,1;x) = C_n x^{1-\alpha}(1-x)^{1-\beta}\frac{d^n}{dx^n}[x^{n+\alpha-1}(1-x)^{n+\beta-1}] \tag{26}$$

$$= C_n F(\alpha+\beta+n-1, -n, \alpha, x), \dagger\ddagger$$

where $F(\alpha, \beta, \gamma, x)$ denotes the hypergeometric series

$$F(\alpha,\beta,\gamma,x) = 1 + \frac{\alpha\cdot\beta}{1\cdot\gamma}x + \frac{\alpha(\alpha+1)\beta(\beta+1)}{1\cdot 2\cdot\gamma(\gamma+1)}x^2 + \cdots. \tag{27}$$

$$a_n = \left[\frac{(\alpha+\beta+n-1)\cdots(\alpha+\beta+2n-2)\Gamma(\alpha+\beta+2n)}{\Gamma(\alpha+n)\Gamma(\beta+n)\Gamma(n+1)}\right]^{1/2}; \tag{28}$$

$$S_n = \frac{n(n+\alpha-1)}{2n+\alpha+\beta-2}. \ddagger$$

$$\Phi_n(\tfrac{1}{2},\beta;0,1;x) \equiv C_n\Phi_{2n}(\beta,\beta\,;-1,1;x^{1/2}) \qquad \text{(see (23))}. \tag{29}$$

Using the asymptotic expressions for Jacobi polynomials (for n very large) derived by Darboux, § we obtain the following asymptotic expressions for $x_{i,n}$-roots of $\Phi_n(\alpha, \beta; 0, 1; x)$, which give a good approximation for $n = 9, 10, \cdots$:

$$x_{i,n} = \left\{\sin^2\left[\frac{\pi(4i+2\alpha-3)}{4(2n+\alpha+\beta-1)}\right]\right\}\cdot\left[1 + \frac{(2\alpha-1)(2\alpha-3)+(2\beta-1)(2\beta-3)}{4(2n+\alpha+\beta-1)(2n-1)}\right] - \frac{(2\alpha-1)(2\alpha-3)}{4(2n+\alpha+\beta-1)(2n-1)} \qquad (i = 1,2,\cdots,n). \tag{30}$$

---

* Darboux, *Mémoire sur l'approximation des fonctions de très-grands nombres et sur une classe étendue de développments en série*, Journal de Mathématiques, (3), vol. 4 (1878), pp. 5–56, 377–416; pp. 377–381.

† Hereafter, C_n stands generally for a constant (different in different formulas) independent of x.

‡ To get the corresponding results for $(-1, 1)$, replace x by $(1+x)/2$, $x_{i,n}$ by $2x_{i,n}-1$.

§ Darboux, loc. cit., p. 44. For the roots of Legendre polynomials in $(-1, 1)$, (30) gives $x_{i,n} = \xi_i[1-(2(4n^2-1))^{-1}]$, which, however, is somewhat inferior to $x_{i,n} = \xi_i[1-(2(2n+1)^2)^{-1}]$, $\xi_i = \cos[\pi(4i-1)[2(2n+1)]^{-1}]$ given by Stieltjes (*Sur les polynomes de Legendre*, Oeuvres, vol. II, pp. 236–252; pp. 243–244).

4. **Existence of formula (A) for any positive integer s. Some properties of $\{n_i\}$.** The results of §3, combined with the conclusion of §2, lead to

THEOREM I. (i) *Formula* (A) *exists for every positive integer* s.

(ii) *The quantities* $(n_i/n)^2$ $(i=1, 2, \cdots, s)$ *are roots of the Jacobi polynomial*

$$\Phi_s(x) \equiv \Phi_s\left(\frac{1}{2}, 2; 0, 1; x\right) = C_s x^{1/2}(1-x)^{-1}\frac{d^s}{dx^s}[x^{s-1/2}(1-x)^{s+1}]$$

$$= C_s F\left(s+\frac{3}{2}, -s, \frac{1}{2}, x\right) = 1 - \frac{(2s+3)s}{1}x$$

$$+ \frac{(2s+3)(2s+5)s(s-1)}{1\cdot 2\cdot 1\cdot 3}x^2 + \cdots{}^* = C_s\Phi_{2s}(2,2;-1,1;x^{1/2}),$$

or, what is equivalent, $\{n_i/n\}$ $(i=1, 2, \cdots, s)$ *coincide with the positive roots* $x_{s+i,2s}$ *of*

$$\Phi_{2s}(2,2;-1,1;x) \equiv C_{2s}F\left(2s+3, -2s, 2, \frac{1+x}{2}\right).$$

Writing $n_{i,s}$ in place of n_i, we state

THEOREM II. (i) $n_{1,s+1} < n_{1,s} < n_{2,s+1} < \cdots < n_{s,s+1} < n_{s,s} < n_{s+1,s+1}$.

(ii) *For* $s\to\infty$, $\lim n_{1,s}=0$, $\lim n_{s,s}=n$.

(iii) *The sequence* $\{n_{i,s}\}$ $(i=1, 2, \cdots, s;\ s=1, 2, \cdots)$ *is everywhere densely distributed in* $(-n, n)$, *i.e., in any sub-interval we find points* $n_{i,s}$, s *being sufficiently large.*

(iv) $n_{1,s}^2+n_{2,s}^2+\cdots+n_{s,s}^2 = n^2\cdot s(2s-1)/(4s+1)$.

(v) $0=\xi_{s+1,2s+1} < n_{1,s}/n < \xi_{s+2,2s+1} < n_{2,s}/n < \cdots < n_{s,s}/n < \xi_{2s+1,2s+1}$, *where* $\xi_{i,n}$ $(i=1, 2, \cdots, n)$ *denote generally the roots of the Legendre polynomial* $P_n(x)$ *corresponding to* $(-1, 1)$.

We need only a proof of (v). The orthogonality and normality properties (22) give readily

$$(1-x^2)\phi_n(2,2;-1,1;x) = A_{n+2}P_{n+2}(x) + A_nP_n(x),$$

(31)

$$A_{n+2} = -\frac{a_n}{p_{n+2}} < 0, \qquad A_n = \frac{p_n}{a_n} > 0 \quad (P_n(x) = p_nx^n + \cdots;\ p_n > 0),$$

* We have also

$$\Phi_s(x) = C_s\begin{vmatrix} \sigma_0 & \sigma_1 \cdots \sigma_s \\ \sigma_1 & \sigma_2 \cdots \sigma_{s+1} \\ \cdots & \cdots \\ \sigma_{s-1} & \sigma_s \cdots \sigma_{2s-1} \\ 1 & x \cdots x^s \end{vmatrix} \qquad \left(\sigma_m = \frac{1}{(2m+1)(2m+3)}\right).$$

and this, combined with the recurrence relation

$$(n+2)P_{n+2}(x) = (2n+3)P_{n+1}(x) - (n+1)P_n(x),$$

gives successively

$$(32)\qquad \begin{aligned} &P_{n+2}(\xi_{i,n+1})P_n(\xi_{i,n+1}) < 0, \quad \phi_n(\xi_{i,n+1})P_n(\xi_{i,n+1}) > 0, \\ &\phi_n(\xi_{i,n+1})\phi_n(\xi_{i+1,n+1}) < 0 \qquad\qquad (i = 1,2, \cdots, n+1). \end{aligned}$$

Hence, the roots of $\phi_n(x)$ are separated by those of $P_{n+1}(x)$, which, in virtue of Theorem I, proves (v).

Note. The roots of Legendre polynomials corresponding to (0, 1) have been computed by Gauss* to 16 decimals, for $n=1, 2, \cdots, 7$, and by Stieltjes (loc. cit.) for $n=9, 10$ (in $(-1, 1)$).

5. **Some properties of the coefficients a_j.** Combining (7), (12), we get

$$(33)\qquad \omega(G_{2s}) = \sum_{j=0}^{s} \omega(\Psi_j)G_{2s}(\alpha_j)/\Psi'(\alpha_j),$$

$$(34)\qquad \omega(\Psi_j G_s) = \omega(\Psi_j)G_s(\alpha_j) \qquad (j = 0,1, \cdots, s),$$

since $G_{2s}(x) \equiv \Psi_j(x)G_s(x) = 0$ $(x=\alpha_i, i\neq j)$, $\Psi'(\alpha_j)G_s(\alpha_j)$ $(x=\alpha_j)$;

$$(35)\qquad b_j = \omega(\Psi_j)/\Psi'(\alpha_j) = \omega(\Psi_j G_s)(\Psi'(\alpha_j)G_s(\alpha_j))^{-1} = \int_0^1 \frac{p_1(y)\Psi(y)G_s(y)dy}{(y-\alpha_j)\Psi'(\alpha_j)G_s(\alpha_j)}$$
$$(p_1(y) = \tfrac{1}{2}y^{-1/2}\,;\ j = 0,1, \cdots, s).$$

Here $G_s(x)$ may be different for different j, and with

$$\frac{G_s(x)}{G_s(\alpha_j)} \equiv \frac{\Psi(x)}{(x-\alpha_j)\Psi'(\alpha_j)},$$

$$(36)\qquad b_j = \frac{a_j}{n} = \int_0^1 p_1(y)\left\{\frac{\Psi(y)}{(y-\alpha_j)\Psi'(\alpha_j)}\right\}^2 dy \qquad (j = 0,1, \cdots, s).$$

THEOREM III. *All coefficients $a_j (j=0, 1, \cdots, s;\ a_0=a)$ are positive.*

This property is of great importance when discussing the convergence of (A). We can also express the a_j in terms of the convergents to the continued fraction (18). We get from (17) rewritten as

---

* Gauss, *Methodus nova integralium valores per approximationem inveniendi*, Werke, vol. III, pp. 105–196; pp. 193–195.

$$\Phi_s(x)\int_0^1 \frac{p(y)dy}{x-y} = P_s(x) + \left(\frac{1}{x^{s+1}}\right) = P_s(x)\int_0^1 p(y)\frac{\Phi_s(x)-\Phi_s(y)}{x-y}dy + \int_0^1 \frac{p(y)\Phi_s(y)dy}{x-y},$$

$$P_s(x) = \int_0^1 p(y)\frac{\Phi_s(x)-\Phi_s(y)}{x-y}dy, \tag{37}$$

and (14) gives, with $\Psi(x) \equiv (x-1)\Phi_s(x)$,

$$b_i = \frac{a_i}{n} = \frac{P_s(\alpha_i)}{(1-\alpha_i)\Phi_s'(\alpha_i)} \qquad (\Phi_s(\alpha_i) = 0\,;\ i = 1,2,\cdots,s). \tag{38}$$

We can also express the a_i in terms of the $\Phi_s(x)$ only, thus *avoiding the computation of the polynomials* $P_s(x)$. Use the properties (ii), (iii), (iv) (§3):

$$\frac{a_i}{n} = \frac{1}{(1-\alpha_i)a_{s-1}^2(p)\Phi_{s-1}(\alpha_i)\Phi_s'(\alpha_i)} = \frac{\alpha_i - c_s}{(1-\alpha_i)a_{s-2}^2(p)\Phi_s'(\alpha_i)\Phi_{s-2}(\alpha_i)}, \tag{39}$$

$$\frac{a_i}{n} = -\frac{(\alpha_i - c_s)H_1(\alpha_i)}{(1-\alpha_i)a_{s-2}^2(p)\Phi_{s-2}(\alpha_i)H_2(\alpha_i)} \quad (\Phi_s(x) = \Phi_s'(x)H_1(x) + H_2(x)), \tag{40}$$

$$\frac{a_i}{n} = \frac{1}{(1-\alpha_i)\sum_{k=0}^{s-1} a_k^2(p)\Phi_k^2(\alpha_i)} = \frac{1}{\sum_{k=0}^{s-1}\phi_k^2(\alpha_i)(1-\alpha_i)} \quad (i = 1,2,\cdots,s). \tag{41}$$

As to the value of $a_0 = a$, it has the following remarkable property:

THEOREM IV. *For any positive integral value of* s, a/n *is rational and* $= [(s+1)(2s+1)]^{-1}$.*

We get readily, for $s=1$, $a/n = 1/6 = 1/(2\cdot 3)$. Assume our statement to be true for $s = 1, 2, \cdots, k$, and prove it holds true also for $s = k+1$. Formula (10): $\omega(G_{2s}) = \int_0^1 p_1(x)G_{2s}(x)dx$, combined with $\int_0^1 p(x)\Phi_s(x)dx = 0 (p(x) = (1-x)p_1(x))$ gives

$$\omega[x\Phi_s(x)] = \omega[\Phi_s(x)]. \tag{42}$$

We get now from the recurrence relation (ii), §3,

$$\Phi_{k+1}(x) = (x - c_{k+1})\Phi_k(x) - \lambda_{k+1}\Phi_{k-1}(x),$$

$$c_k = \frac{8k^2 - 4k - 3}{(4k+1)(4k-3)},\quad \lambda_{k+1} = \frac{(k^2 - \frac{1}{4})k(k+1)}{(2k-\frac{1}{2})(2k+\frac{1}{2})^2(2k+\frac{3}{2})} \tag{43}$$

* In this remarkable property of a/n lies the great practical value of formula (A), since adding two more ordinates (at $x = \pm n$) contributes greatly to the accuracy of (A), while requiring very little computation.

(see (28) and (iii), §3; $k \geqq 1$):

$$(44) \qquad \frac{\omega(\Phi_{k+1})}{\Phi_{k+1}(1)} = (1 - c_{k+1})\frac{\omega(\Phi_k)}{\Phi_k(1)}\,\frac{\Phi_k(1)}{\Phi_{k+1}(1)} - \lambda_{k+1}\frac{\omega(\Phi_{k-1})}{\Phi_{k-1}(1)}\,\frac{\Phi_{k-1}(1)}{\Phi_{k+1}(1)}.$$

We obtain finally from (44), substituting therein (Darboux, loc. cit.)

$$(45) \quad \Phi_s(1) = \Phi_s(\alpha,\beta;0,1;1) = \frac{\beta(\beta+1)\cdots(\beta+n-1)}{(\alpha+\beta+n-1)\cdots(\alpha+\beta+2n-2)} \qquad (\text{with } \alpha = \tfrac{1}{2},\ \beta = 2),$$

a/n (for $s = k+1) = 1/[(k+2)\ (2k+3)]$, which proves our statement.

6. **Generalization of formula (A).** Choose, in (a, b), $n+1$ arbitrary points $x_0, x_1, \cdots, x_n$. With $p(x)$ defined on (a, b) and such that $\int_a^b p(x)x^i dx$ exists $(i = 0, 1, 2, \cdots)$, we write the following formula of mechanical quadratures (making use of the Lagrange interpolation formula):

$$(46) \qquad \int_a^b p(x)G_n(x)dx = \sum_{i=0}^{n} A_i G_n(x_i),$$

$$A_i = \int_a^b \frac{p(x)\phi(x)dx}{(x - x_i)\phi'(x_i)}, \qquad \phi(x) = \prod_{i=0}^{n}(x - x_i).$$

Moreover, there exists a remarkable choice of the x_i, which we state* in

THEOREM V. *If $p(x)$ be a c-function defined on a finite interval (a, b), we can always choose $2s+2$ points $x_0 = a < x_1 < x_2 < \cdots < x_{2s} < x_{2s+1} = b$ so that* (46) *holds true for $G_{4s+1}(x)$. The points $x_1, \cdots, x_{2s}$ are roots of the polynomial $\Phi_{2s}[p(x)\ (x-a)\ (b-x);\ a, b;\ x]$ of the family of orthogonal Tchebycheff polynomials corresponding to the interval (a, b), with the "characteristic function" $p(x)\ (x-a)\ (b-x)$.*

For any $f(x)$ defined on (a, b) we thus write:*

$$(B) \qquad \int_a^b p(x)f(x)dx = \sum_{i=0}^{2s+1} A_i f(x_i) + R_{2s+2}(f) \qquad (x_0 = a,\ x_{2s+1} = b),$$

$$A_i = \int_a^b \frac{p(x)\Phi(x)dx}{(x - x_i)\phi'(x_i)},$$

$$\Phi(x) = (x-a)(x-b)\Phi_{2s}[p(x)(x-a)(b-x);\, a, b;\, x].$$

For the coefficients A_i we can get expressions quite similar to those given above for the a_j in (A) (formulas (12)–(14), (39)–(41)). One interesting con-

* A. Markoff, *Differenzenrechnung*, Leipzig, Teubner, 1896, pp. 69, 80–87. Omitting continuity of $f^{(4s+2)}(x)$ in (a, b), we may replace in (47) $f^{(4s+2)}(\xi)$ by a quantity intermediate between the upper and lower bounds of $f^{(4s+2)}(x)$ in (a, b).

clusion is the following: *if a, b are rational, and all the "moments" of $p(x)$ are likewise rational, then the coefficients A_0, A_{2s+1}, corresponding to the end points of the interval, are also rational.*

If $f^{(4s+2)}(x)$ be continuous in (a, b), we get for the remainder in (B)*

$$R_{2s+2}(f) = -\frac{1}{\Gamma(4s+3)} f^{(4s+2)}(\xi) \int_a^b p(x)(x-a)(b-x)\Phi_{2s}^2(x)dx$$

$$= -\frac{f^{(4s+2)}(\xi)}{\Gamma(4s+3)} \frac{1}{a_{2s}^2 [p(x)(x-a)(b-x)]} \qquad (\xi \text{ in } (a,b)). \tag{47}$$

Making use of the lower bound for $a_n(p)$ given by the writer,† we get

$$|R_{2s+2}(f)| \leqq \frac{M_{4s+2}}{\Gamma(4s+3)} \cdot 4\left(\frac{b-a}{4}\right)^{4s} \cdot \int_a^b p(x)(x-a)(b-x)dx,$$

$$M_{4s+2} = \max |f^{(4s+2)}(x)| \text{ in } (a,b). \tag{48}$$

Considerations similar to those employed above in the proof of Theorem II, (v), show that the roots of $\Phi_n[p(x)(x-a)(b-x); x]$ are separated by those of $\Phi_{n+1}[p(x); a, b; x]$.

THEOREM VI. *The $2s$ points $x_1, \cdots, x_{2s}$ employed in* (B) *satisfy the inequalities* $x_{1,2s+1} < x_1 < x_{2,2s+1} < \cdots < x_{2s,2s+1} < x_{2s} < x_{2s+1,2s+1}$, *where* $x_{i,2s+1}$ $(i=1, \cdots, 2s+1)$ *are roots of* $\Phi_{2s+1}[p(x); a, b; x]$.

For the coefficients A_i in (B) we have the important Tchebycheff inequalities‡

$$A_i + A_{i+1} + \cdots + A_k > \int_{x_i}^{x_k} p(x)dx > A_{i+1} + \cdots + A_{k-1} \tag{49}$$

$$(i, k = 0, 1, \cdots, 2s+1),$$

$$0 < A_i < \int_{x_{i-2}}^{x_{i+2}} p(x)dx \tag{50}$$

$$(i = 0, 1, \cdots, 2s+1;\ x_{-2} = x_{-1} = a,\ x_{2s+2} = x_{2s+3} = b),$$

which, combined with (vii), (viii) of §3, leads to

THEOREM VII. *If* $\int_\alpha^\beta p(x)dx > 0$ $(a \leqq \alpha < \beta \leqq b)$, *then* $\lim_{s\to\infty} A_i = 0$ $(i=0, 1, \cdots, 2s+1)$.

* See preceding footnote.

† J. Shohat, these Transactions, vol. 29 (1927), pp. 569–583; p. 575.

‡ (a) A. Markoff, *Démonstration de certaines inégalités de M. Tchebycheff*, Mathematische Annalen, vol. 24 (1884), pp. 172–180, pp. 172–173; (b) Stieltjes, *Quelques recherches sur la théorie des quadratures dites mécaniques*, Oeuvres, vol. I, pp. 377–394; pp. 385, 387, 392–394.

7. **Convergence of formula (B).** Consider, first, the case of $f(x)$ *continuous in* (a, b). Let $\Pi_{4s+1}(x)$ denote the polynomial of degree $\leqq 4s+1$, of the best approximation (in the sense of Tchebycheff) to $f(x)$ on (a, b), and let

$$(51)\quad E_{4s+1}(f)\ (\text{"best approximation"}) = \max. \left| f(x) - \Pi_{4s+1}(x) \right| \text{ for } a \leqq x \leqq b.$$

Apply (B) to $f(x)$ and $\Pi_{4s+1}(x)$, and use (50) and (take $f(x) \equiv 1$ in (B)):

$$(52)\quad \sum_{i=0}^{2s+1} A_i = \int_a^b p(x)dx,$$

$$R_{2s+2}(f) = \int_a^b p(x)[f(x) - \Pi_{4s+1}(x)] + \sum_{i=0}^{2s+1} A_i[\Pi_{4s+1}(x) - f(x_i)],$$

$$(53)\quad |\, R_{2s+2}(f)\,| \leqq 2E_{4s+1}(f) \int_a^b p\,(x)\,dx.$$

By Weierstrass's theorem, $\lim_{s\to\infty} E_{4s+1}(f) = 0$; hence, for $s \to \infty$, formula (B) converges for any continuous function.

Furthermore, the inequalities (49), (50) enable us to follow the elegant analysis of Stieltjes* and state

THEOREM VIII. *For* $s \to \infty$, *formula* (B) *converges, i.e.* $\lim_{s\to\infty} \sum_{i=0}^{2s+1} A_i f(x_i) = \int_a^b p(x)f(x)dx$, *for any* $f(x)$ *bounded in* (a, b), *for which the right-hand* (R) *integral exists.*

Notes. (i) Employing Riemann-Stieltjes integrals, the writer was able to prove† that *Theorems* VII, VIII *are valid even without the condition* $\int_\alpha^\beta p(x)dx > 0$ $(a \leqq \alpha < \beta \leqq b)$.

(ii) Theorem VIII could be established without the use of Tchebycheff inequalities, following an elementary analysis employed by Stekloff‡ for $p(x) \equiv 1$, which, however, would require considerable supplements for our general $p(x)$.

8. **Accuracy of formula (B) for continuous functions.** Formula (53) gives, for any s, the order of magnitude of $R_{2s+2}(f)$, using therein the known order (with respect to s) of the best approximation $E_{4s+1}(f)$, $f(x)$ having cer-

---

* See preceding foot note, reference (b).

† J. Chokhate, *Sur la convergence des quadratures mécaniques dans un intervalle infini. Applications au problème des moments, au calcul des probabilités*, Comptes Rendus, vol. 186, pp. 344–346.

‡ W. Stekloff, *On the approximate evaluation of definite integrals by means of mechanical quadratures. I. Convergence of mechanical quadratures formulas* (in Russian), Bulletin of the Russian Academy of Sciences, 1916, pp. 170–186; pp. 176–178.

tain prescribed continuity properties. The following cases make use of certain results in this domain due to D. Jackson* and C. de la Vallée-Poussin† and are well adapted for numerical computation.

$$(54)\quad \begin{matrix}(\alpha)\,|f(x_2)-f(x_1)|\leqq\lambda|x_1-x_2|\ ;\ (\beta)\,|f'(x_2)-f'(x_1)|\leqq\lambda_1|x_1-x_2|\ ;\\ (\gamma)\,|f(x_2)-f(x_1)|\leqq\omega(\delta)\ ;\qquad (\delta)\,|f'(x_2)-f'(x_1)|\leqq\omega_1(\delta)\end{matrix}$$
$$(a\leqq x_1,\ x_2\leqq b\ ;\ \lambda,\lambda_1=\text{const.}\ ;\ |x_1-x_2|\leqq\delta).$$

$$(55)\quad \begin{matrix}(\alpha)\,|R_{2s+2}(f)|<3\lambda\frac{b-a}{(4s+1)}Q^2\ ;\ (\beta)\,|R_{2s+2}(f)|<20\lambda_1\left(\frac{b-a}{4s+1}\right)^2Q^2\ ;^*\\ (\gamma)\,|R_{2s+2}(f)|<6\omega\left(\frac{b-a}{4s+2}\right)Q^2\ ;\ (\delta)\,|R_{2s+2}(f)|<5\frac{b-a}{2s+1}\omega_1\left(\frac{b-a}{4s+2}\right)Q^2\ ;^\dagger\end{matrix}$$
$$\left(Q^2=\int_a^b p(x)dx\right).$$

9. **Symmetric formulas of mechanical quadratures.** We get in (B), with

$$(56)\qquad a=-b,\qquad p(x)\equiv p(-x)\quad\text{in}\quad(-b,b),$$

$$(57)\qquad x_i+x_{2s+1-i}=0,\ A_i=A_{2s+1-i}\qquad(i=0,1,\cdots,2s-1)$$

(see (ix), §3), and we may call such a formula a "*symmetric*" *formula of mechanical quadratures*. Rewrite now formula (A) as follows:

$$(A)\qquad \int_{-n}^{n}u_x dx=a(u_{-n}+u_n)+\sum_{i=1}^{s}a_i(u_{-n_i}+u_{n_i})+\rho_s(u),$$

$$\left[(A')\qquad \int_0^{2n}u_x dx=a(u_0+u_{2n})+\sum_{i=1}^{s}a_i(u_{n-n_i}+u_{n+n_i})+\rho_s(u)\right],$$

and we arrive at the following conclusions:

THEOREM IX. (i) *Formula* (A) *is a special case of "symmetric" formulas of mechanical quadratures* (B), *with* $p(x)\equiv 1$, $(a, b)=(-1, 1)$, *if we replace in* (B) x *by* x/n (*with* $f(x/n)\equiv u_x$).

(ii) $\lim_{s\to\infty}a=\lim_{s\to\infty}a_i=0$ $(i=1,\cdots,s)$.

(iii) *Formula* (A) *converges, for* $s\to\infty$, *for any* $u(x)$ *bounded and* (*R*)-*integrable in* $(-n, n)$.

(iv) *If* $u^{(4s+2)}(x)$ *is continuous in* $(-n, n)$, *then* (*see* (47), (48), *also third footnote on page* 458)

* D. Jackson, *On approximation by trigonometric sums and polynomials*, these Transactions, vol. 13 (1912), pp. 491–515; pp. 492, 496, 510, 512.

† C. de la Vallée-Poussin, *Leçons sur l'Approximation des Fonctions d'une Variable Réelle*, Paris, 1919, pp. 66–67.

$$(58)\quad \rho_s(u) = -n\frac{2^{4s+4}(s+1)(2s+1)^3\Gamma^4(2s+1)}{(4s+3)\Gamma^3(4s+3)}u^{(4s+2)}(\xi) \quad (\xi \text{ in } (-n,n)).$$

(v) *If $f(x)$ satisfies one of the conditions* (54α-δ), *then we have the corresponding inequalities*

$$(59)\quad \begin{array}{ll}(\alpha)\ |\rho_s(u)|<12n\lambda/(4s+1); & (\beta)\ |\rho_s(u)|<160n\lambda_1/(4s+1)^2;\\ (\gamma)\ |\rho_s(u)|<12n\omega(1/(2s+1)); & (\delta)\ |\rho_s(u)|<(20n/(2s+1))\omega(1/(2s+1)).\end{array}$$

10. Extremal properties of the a_i in (A). Apply (B), with $p(x)=1$, $(a, b)=(-1, 1)$, to the polynomial $(1-x^2)\,G_{s-1}^2(x^2)$:

$$(60)\quad \begin{aligned}&\int_0^1(1-x^2)G_s^2{}_{-1}(x^2)dx = \sum_{i=0}^{s}A_i(1-x_i^2)G_{s-1}^2(x_i^2),\\ &\int_0^1\frac{1}{2}x^{-1/2}(1-x)G_s^2{}_{-1}(x)dx = \sum_{k=0}^{s}A_k(1-\alpha_k)G_{s-1}^2(\alpha_k) \geqq A_i(1-\alpha_i)G_{s-1}^2(\alpha_i)\\ &\qquad\qquad (s \geqq i > 0;\ \alpha_k = x_k^2,\ k = 0,1,2,\cdots,\ s).\end{aligned}$$

Consider now (60) for all polynomials $(1-x)\,G_{s-1}(x)$ such that

$$G_{s-1}(\alpha_i) = \frac{1}{1-\alpha_i} \qquad (i>0).$$

Then,

$$\int_0^1\frac{1}{2}x^{-1/2}(1-x)G_s^2{}_{-1}(x)dx \geqq A_i \qquad (i = 1,2,\cdots,s),$$

$$(61)\quad A_i = \min\int_0^1\frac{1}{2}x^{-1/2}(1-x)G_s^2{}_{-1}(x)dx \qquad (G_{s-1}(\alpha_i)(1-\alpha_i) = 1).$$

We know that A_i, α_i in (60) are the same, respectively, as a_i/n, $(n_i/n)^2$ in (A) $(i=1,\cdots,s)$. Thus we get from (61), using a result previously established by the writer,*

$$(62)\quad A_i = \frac{1}{(1-\alpha_i)\sum_{k=0}^{s-1}\phi_k^2[\frac{1}{2}x^{-1/2}(1-x);\,0,1;\,\alpha_i]} \qquad (i = 1,2,\cdots,s),$$

which is the same formula (41) derived above in an entirely different way. The extremal properties (61) are of great importance in discussing formulas of mechanical quadratures in general. They enable us, for example, to find upper bounds for the coefficients A_i in the formulas under consideration, to

* These Transactions, vol. 29 (1927) p. 573.

prove the *uniform convergence to zero of the A's in formulas of type* (B),* and so on. All these considerations will be developed elsewhere.

11. **Numerical values of the constants in formula (A) for some values of** s. First, construct the hypergeometric polynomial $\Phi_s(x) = C_s F_s(s+\frac{3}{2}, -s, \frac{1}{2}, x)$ $(= x^s + \cdots)$ the roots of which give the quantities $(n_i/n)^2$ $(i = 1, \cdots, s)$; then find the λ's and c's by means of ((43), (iii), §3) (or dividing $\Phi_s(x)$ by $\Phi_{s-1}(x)$); formulas (39) or (40) yield now the values of a_i/n $(i = 1, \cdots, s)$††, a/n being equal to $1/[(s+1)(2s+1)]$. For $s = 5, 6, \cdots$, we can find the quantities n_i/n as the positive roots of the polynomial $F(2s+3, -2s, ?, (1+x)/2)$, using for those roots their approximate expressions (30) (with $\alpha = \beta = 2$).

$$s = 2: \quad \Phi_2(x) = x^2 - \frac{2}{3}x + \frac{1}{21}; \quad \frac{n_1}{n} = 0.28522, \quad \frac{n_2}{n} = 0.76506;$$

$$\frac{a}{n} = \frac{1}{15}; \quad \frac{a_1}{n} = 0.55486; \quad \frac{a_2}{n} = 0.37847;$$

$$\rho_2(u) = -n \cdot \frac{2^{12} \cdot 3 \cdot 5^3 \cdot (4!)^4}{11 \cdot (10!)} u^{(10)}(\xi) \qquad (\xi \text{ in } (-n, n));$$

$$|\rho_2(u)| < \frac{4}{3}n\lambda, \quad \frac{160}{81}n\lambda_1, \quad 12n\omega\left(\frac{1}{5}\right), \quad 4n\omega_1\left(\frac{1}{5}\right) \text{ (conditions (54}\alpha\text{-}\delta\text{))}.$$

$$s = 3: \quad \Phi_3(x) = x^3 - \frac{3 \cdot 5}{13}x^2 + \frac{5 \cdot 9}{11 \cdot 13}x - \frac{5}{3 \cdot 11 \cdot 13};$$

$$\frac{n_1}{n} = 0.20930, \quad \frac{n_2}{n} = 0.59170, \quad \frac{n_3}{n} = 0.87174;$$

$$\frac{a}{n} = \frac{1}{28}, \quad \frac{a_1}{n} = 0.41245, \quad \frac{a_2}{n} = 0.34111, \quad \frac{a_3}{n} = 0.21069;$$

$$\rho_3(u) = -n \cdot \frac{2^{18} \cdot 7^3 \cdot (6!)^4}{15 \cdot (14!)^3} u^{(14)}(\xi) \qquad (\xi \text{ in } (-n, n));$$

$$|\rho_3(u)| < \frac{12n\lambda}{13}, \quad \frac{160n\lambda}{169}, \quad 12n\omega\left(\frac{1}{7}\right), \quad \frac{20}{7}n\omega_1\left(\frac{1}{7}\right) \text{ (conditions (54}\alpha\text{-}\delta\text{))}.$$

---

* Cf. J. Chokhate, Comptes Rendus, loc. cit.

† When using (39), (40), add to (28) (with $n = 2s$, $\alpha = \frac{1}{2}$, $\beta = 2/2^{1/2}$, as factor, due to $p(x) = \frac{1}{2}x^{-1/2} \cdot (1-x)$).

‡ Check: $\frac{a}{n} + \sum_{i=1}^{s} \frac{a_i}{n} = 1; \qquad \sum_{i=1}^{s} \left(\frac{n_i}{n}\right)^2 = s(2s-1)/(4s+1).$

$$s = 4: \quad \Phi_4(x) = x^4 - \frac{4\cdot 7}{17}x^3 + \frac{2\cdot 7}{17}x^2 - \frac{4\cdot 7}{13\cdot 17}x + \frac{7}{11\cdot 13\cdot 17};$$

$$\frac{n_1}{n} = 0.16528, \quad \frac{n_2}{n} = 0.47792, \quad \frac{n_3}{n} = 0.73877, \quad \frac{n_4}{n} = 0.91953.*$$

* I am indebted to Professor H. C. Carver for this computation.

UNIVERSITY OF MICHIGAN,
ANN ARBOR, MICH.

GENERALIZED VANDERMONDE DETERMINANTS*

BY

E. R. HEINEMAN

INTRODUCTION

The Vandermonde determinant, usually written in this way:

$$\begin{vmatrix} 1 & \cdots & 1 \\ a_1 & \cdots & a_n \\ \cdots & \cdots & \cdots \\ a_1^{n-1} & \cdots & a_n^{n-1} \end{vmatrix},$$

is an alternant and is equal to the difference-product of its variables. Instead of keeping the indices of the Vandermonde determinant fixed at $0, 1, 2, \cdots, n-1$, let us take any positive integers. This yields another alternant which we shall call a generalized Vandermonde determinant. This determinant may be expressed thus:

$$\begin{vmatrix} a_1^{t_1} & \cdots & a_n^{t_1} \\ a_1^{t_2} & \cdots & a_n^{t_2} \\ \cdots & \cdots & \cdots \\ a_1^{t_n} & \cdots & a_n^{t_n} \end{vmatrix}$$

in which $t_1 < t_2 < \cdots < t_n$. Since this alternant vanishes on equating any two of its variables, we know that every generalized Vandermonde determinant is divisible by the difference-product of its variables and hence by the Vandermonde determinant of these variables. Let us now consider the quotient of any generalized Vandermonde determinant by its Vandermonde determinant. Since both change sign under a transposition, their quotient will remain unchanged, putting it into the class of symmetric functions. A general formula for finding this symmetric function has been the goal of much research in the last half century.

Thus far two ways of treating this subject have been introduced. One of these involves the determination of a general method for finding the quotient of a generalized Vandermonde determinant by the Vandermonde determinant of its variables in terms of symmetric functions. This method

* Presented to the Society, September 9, 1927; received by the editors in February, 1929.

has been used by W. Woolsey Johnson* and, in an entirely different manner, by Thomas Muir.† The second method of approach, used by both Muir‡ and E. D. Roe, Jr.,§ consists of the multiplication of the Vandermonde determinant by a symmetric function and the expression of this product as an integral rational function of generalized Vandermonde determinants in the same variables.

The object of this paper is to develop a new method of treating this problem, consisting in first expressing every generalized Vandermonde determinant as an integral rational function of certain special Vandermonde determinants. This leads to a solution of the problem mentioned above; and it gives, in conjunction with a theorem of Muir's, a direct method for expressing an integral symmetric function of n variables in terms of the elementary symmetric functions. In this treatment of the subject, the determinant

$$\begin{vmatrix} 1 & \cdots & 1 \\ a_1 & \cdots & a_n \\ \cdots & \cdots & \cdots \\ a_1^{n-1} & \cdots & a_n^{n-1} \end{vmatrix} = \prod_{\substack{i=2 \\ i>j}}^{n}(a_i - a_j) \quad (j=1,2,\cdots,n-1)$$

will be called the principal Vandermondian and will be represented by the notation $|n-1, n-2, \cdots, 1, 0|$. The matrix formed from this determinant by adding the nth power of the variables as an extra row will be called the Vandermonde matrix and will be represented by the symbol $\|n, n-1, \cdots, 1, 0\|$. And the determinant formed from this matrix by omitting the $(n-k+1)$th row, or the $(k+1)$th term in the symbol, will be represented by $|n, n-1, \cdots, n-k+1, n-k-1, \cdots, 1, 0|$ or briefly by V_{nk}, in which the first subscript, n, denotes the number of variables, and the second subscript, k, indicates that the $(k+1)$th term has been omitted from the symbol for the Vandermonde matrix, $\|n, n-1, \cdots, 1, 0\|$.¶ The special case $k=0$ gives the principal Vandermondian V_{n0}. The determinants V_{nk}, where $k \neq 0$, will be called secondary Vandermondians. The first section of this paper is concerned with expressing any generalized Vandermonde determinant in terms of the principal and secondary Vandermondians.

---

* American Journal of Mathematics, vol. 7, pp. 345–352, 380–388; Quarterly Journal of Mathematics, vol. 21, pp. 217–224.

† Proceedings of the Royal Society of Edinburgh, vol. 14, pp. 433–445.

‡ Proceedings of the Royal Society of Edinburgh, vol. 22, pp. 539–542.

§ American Journal of Mathematics, vol. 25, pp. 97–106.

¶ Where there is no danger of confusion, the subscript n will be omitted from the symbol V_{nk}.

1. Secondary Vandermondians and elementary symmetric functions. Notations

Since the quotient of any generalized Vandermonde determinant by the principal Vandermondian is always a symmetric function, we can express it rationally in terms of elementary symmetric functions. We will now show that every elementary symmetric function of $a_1, \cdots, a_n$ is expressible rationally in terms of the principal and secondary Vandermondians in these variables. Consider the equation

$$p_0x^n + p_1x^{n-1} + \cdots + p_{n-1}x + p_n = 0, \tag{1}$$

whose roots are $a_1, a_2, \cdots, a_n$. We then have

$$\frac{p_i}{p_0} = (-1)^i \sum a_1a_2 \cdots a_i \qquad (i = 1,2, \cdots, n). \tag{2}$$

By substituting each of the roots in equation (1), we obtain n identities:

$$\sum_{j=0}^{n} p_j a_i^{n-j} = 0 \qquad (i = 1,2, \cdots, n). \tag{3}$$

We look upon these identities as stating that $p_0, p_1, \cdots, p_n$ satisfy a system of n equations in the p's. The matrix of this system of n equations in $n+1$ unknowns is the Vandermonde matrix $\|n, n-1, \cdots, 1, 0\|$. Hence we have

$$\frac{p_i}{p_0} = (-1)^i \frac{V_{ni}}{V_{n0}} \qquad (i = 1,2, \cdots, n). \tag{4}$$

Equations (2) and (4) yield

$$\frac{V_{ni}}{V_{n0}} = \sum a_1a_2 \cdots a_i \qquad (i = 1,2, \cdots, n). \tag{5}$$

We may therefore state

Theorem I. *The elementary symmetric function* $E_i \equiv \sum a_1a_2 \cdots a_i$ *of the* n *variables* $a_1, a_2, \cdots, a_n$ *is equal to the quotient of the secondary Vandermondian* V_{ni} *by the principal Vandermondian* V_{n0}.

From this theorem follows

Corollary I. $p_j = (-1)^j V_{nj}\rho$, *where* ρ *is a proportionality factor.*

Before proceeding farther, we will lay down the following definitions:

Definition I. D_l^n will represent the following lth order determinant, formed from principal and secondary Vandermondians in n variables:

$$D_l^n \equiv \begin{vmatrix} V_1 & V_2 & V_3 \cdots & V_n & 0 & \cdots & 0 \\ V_0 & V_1 & V_2 \cdots & 0 & V_n & \cdots & 0 \\ 0 & V_0 & V_1 \cdots & \cdots & V_n & 0 \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & \cdots & \cdots & 0 & V_0 & V_1 \quad V_2 \\ 0 & \cdots & \cdots & \cdots & \cdots & 0 & V_0 \quad V_1 \end{vmatrix}_l,$$

where l may be less than, equal to, or greater than n.

DEFINITION II. $D_l^n\{t_1, t_2, \cdots, t_{s-1}, t_s\}$, where $l \geqq t_1 \geqq t_2 \geqq \cdots \geqq t_{s-1} \geqq t_s$, is defined to be the determinant obtainable from D_l^n by increasing the subscripts in the first t_s rows by s, those in the next $t_{s-1}-t_s$ rows by $s-1, \cdots$, those in the next t_1-t_2 rows by 1. In this connection V_m is defined to be 0 if m is greater than n; and a zero which precedes a V_0 should be replaced by V_0 whenever the subscripts in the row in which it stands are increased by 1. For example

$$D_4^5\{3,2,1,1\} \equiv \begin{vmatrix} V_5 & 0 & 0 & 0 \\ V_2 & V_3 & V_4 & V_5 \\ V_0 & V_1 & V_2 & V_3 \\ 0 & 0 & V_0 & V_1 \end{vmatrix}.$$

2. THE DETERMINANTS $|k, n-2, n-3, \cdots, 1, 0|$

We shall now attempt to develop a general formula for the generalized Vandermonde determinant $|k, n-2, n-3, \cdots, 1, 0|$, where $k > n-2$. To establish this general formula we will use mathematical induction. Setting $k = n-1$ yields the principal Vandermondian V_{n0}. Also, if $k = n$, we have the secondary Vandermondian V_{n1}. In comparing $|k+1, n-2, \cdots, 1, 0|$, where $k > n$, with $|k, n-2, \cdots, 1, 0|$, we shall consider the n systems of $k-n+2$ equations each:

$$\begin{aligned} p_0 a_i^{k+1} + p_1 a_i^k + \cdots + p_n a_i^{k-n+1} &= 0, \\ p_0 a_i^k + \cdots + p_{n-1} a_i^{k-n+1} + p_n a_i^{k-n} &= 0, \\ \cdots\cdots\cdots\

$$\begin{vmatrix}
p_2 & p_3 & \cdots & p_n & 0 & \cdots & \cdots & \cdots & \cdots & 0 & p_0 a_i^{k+1} + p_1 a_i^{k} \\
p_1 & p_2 & \cdots & p_n & 0 & \cdots & \cdots & \cdots & \cdots & 0 & p_0 a_i^{k} \\
p_0 & p_1 & \cdots & p_n & 0 & \cdots & \cdots & \cdots & \cdots & 0 & 0 \\
0 & p_0 & \cdots & p_n & 0 & \cdots & \cdots & \cdots & \cdots & 0 & 0 \\
\cdot & \cdot & \cdots & \cdot & \cdot & \cdots & \cdots & \cdots & \cdots & \cdot & \cdot \\
0 & \cdots & \cdots & 0 & p_0 & \cdots & \cdots & \cdots & \cdots & p_n & 0 \\
0 & \cdots & \cdots & 0 & p_0 & \cdots & \cdots & \cdots & \cdots & p_{n-1} & p_n a_i^{n-2} \\
0 & \cdots & \cdots & 0 & p_0 & \cdots & \cdots & \cdots & \cdots & p_{n-2} & p_{n-1} a_i^{n-2} + p_n a_i^{n-3} \\
\cdot & \cdot & \cdots & \cdot & \cdot & \cdots & \cdots & \cdots & \cdots & \cdot & \cdot \\
0 & \cdots & \cdots & 0 & p_0 & p_1 & p_2 & \cdots & \cdots & p_3 a_i^{n-2} + \cdots + p_n a_i & \\
0 & \cdots & \cdots & 0 & p_0 & p_1 & p_2 & \cdots & \cdots & p_2 a_i^{n-2} + \cdots + p_{n-1} a_i + p_n &
\end{vmatrix}_{k-n+2} = 0.$$

Multiplying the first row by $(-p_0)$ and adding to it p_1 times the second row and developing according to the elements of the last column, we get

$$a_i^{k+1} p_0^2 \begin{vmatrix}
p_1 & p_2 & \cdots & p_n & 0 & \cdots & 0 \\
p_0 & p_1 & \cdots & p_n & 0 & \cdots & 0 \\
\cdot & \cdot & \cdots & \cdot & \cdot & \cdots & \cdot \\
0 & \cdots & \cdots & 0 & p_0 & p_1 &
\end{vmatrix}_{k-n+1}$$

$$+ a_i^k p_0 \begin{vmatrix}
p_1^2 - p_0 p_2 & p_1 p_2 - p_0 p_3 & \cdots & p_1 p_{n-1} - p_0 p_n & p_1 p_n & 0 & \cdots & 0 \\
p_0 & p_1 & \cdots & p_{n-2} & p_{n-1} & p_n & 0 & \cdots 0 \\
0 & p_0 & \cdots & p_{n-3} & p_{n-2} & p_{n-1} & p_n & 0 \cdots 0 \\
\cdot & \cdot & \cdots & \cdot & \cdot & \cdot & \cdot & \cdot \\
0 & \cdot & \cdots & \cdot & \cdot & \cdot & 0 & p_0 p_1
\end{vmatrix}_{k-n+1}$$

$$+ M_2 a_i^{n-2} + M_3 a_i^{n-3} + \cdots + M_{n-1} a_i + M_n = 0,$$

where the M's represent polynomials in the p's. The second determinant, which, multiplied by p_0, is the coefficient of a_i^k in the above-mentioned equation, may be looked upon as the sum of two determinants or as a single determinant of order one higher, giving us

$$a_i^{k+1} p_0^2 \begin{vmatrix}
p_1 & p_2 & \cdots & p_n & 0 & \cdots & 0 \\
p_0 & p_1 & \cdots & p_n & 0 & \cdots & 0 \\
0 & p_0 & \cdots & p_n & 0 & \cdots & 0 \\
\cdot & \cdot & \cdots & \cdot & \cdot & \cdots & \cdot \\
0 & \cdots & \cdots & 0 & p_0 & p_1 &
\end{vmatrix}_{k-n+1}$$

$$+ a_i^k p_0 \begin{vmatrix} p_1 & p_2 \cdots p_n & 0 & \cdots 0 \\ p_0 & p_1 \cdots \cdots & p_n & 0 \cdots 0 \\ 0 & p_0 \cdots \cdots & p_n & 0 \cdots 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots \cdots & 0 & p_0 \; p_1 \end{vmatrix}_{k-n+2} + M_2 a_i^{n-2} + \cdots + M_n = 0.$$

Using the corollary to Theorem I to change the p's to the V's, multiplying the odd rows and the even columns of the two determinants through by -1, dividing through by V_0 and by the proper power of the proportionality factor ρ, and introducing the symbol D_r^n, we obtain n equations which may be considered as homogeneous equations in D_{k-n+2}^n, D_{k-n+1}^n, and $n-1$ other variables which differ from the M's by the factor V_0, multiplied by some power of ρ. Solving these equations for the ratio of D_{k-n+2}^n and D_{k-n+1}^n, we find

$$V_0 | \, k+1, n-2, \cdots, 1,0 | \; : \; | \, k, n-2, \cdots, 1,0 | = D_{k-n+2}^n : D_{k-n+1}^n.$$

Writing this relation for $k = n-1, n, n+1, \cdots, k-1$, we get

$$V_0 V_1 : V_0 = D_1^n : D_0^n = V_1 : 1,$$

$$V_0 | \, n+1, n-2, \cdots, 1,0 | : V_1 = D_2^n : D_1^n,$$

$$V_0 | \, n+2, n-2, \cdots, 1,0 | \; : \; | \, n+1, n-2, \cdots, 1,0 | = D_3^n : D_2^n,$$

$$\cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots$$

$$V_0 | \, k, n-2, \cdots, 1,0 | \; : \; | \, k-1, n-2, \cdots, 1,0 | = D_{k-n+1}^n : D_{k-n}^n.$$

Multiplying the terms of the first ratios in each of these equations by 1, V_0, V_0^2, $\cdots$, V_0^{k-n}, we get the continued proportion

$$V_0^{k-n+1} | \, k, n-2, \cdots, 1,0 | : V_0^{k-n} | \, k-1, n-2, \cdots, 1,0 | : \cdots$$

$$: V_0^2 | \, n+1, n-2, \cdots, 1,0 | : V_0 V_1 : V_0 = D_{k-n+1}^n : D_{k-n}^n : \cdots : D_2^n : D_1^n : 1.$$

Hence we find that

$$V_0^{k-n} | \, k, n-2, \cdots, 1,0 | = D_{k-n+1}^n.$$

We have therefore proved the following theorem:

THEOREM II. *The generalized Vandermonde determinant* $|k, n-2, \cdots, 1, 0|$ *multiplied by the* $(k-n)$*th power of the principal Vandermondian is equal to the determinant* D_{k-n+1}^n (*see Definition* I).

3. The determinants $|k, j, n-3, \cdots, 1, 0|$

Our next task is to develop a formula for the determinant in which the two highest exponents of the variables are allowed to vary. We shall attempt to compare $|k, j, n-3, \cdots, 1, 0|$, where $j>n-2$, with $|k, n-2, \cdots, 1, 0|$. To do this we consider n systems of $k-n+1$ equations each:

$$\begin{aligned} p_0a_i^k + p_1a_i^{k-1} + \cdots + p_{k-j}a_i^j + \cdots + p_na_i^{k-n} &= 0, \\ p_0a_i^{k-1} + \cdots + p_{k-j-1}a_i^j + \cdots + p_na_i^{k-n+1} &= 0, \\ \cdots\cdots\cdots\

$$+\begin{vmatrix} p_1 \cdots p_{k-j-1}\ p_{k-j+1} \cdots p_n\ 0 \cdots 0 \\ p_0 \cdots p_{k-j-2}\ p_{k-j} \cdots p_n\ 0 \cdots 0 \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ 0 \cdots 0\ p_0 \cdots p_{n-1}\ p_n \\ 0 \cdots 0\ p_0 \cdots p_{n-2}\ p_{n-1} \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ 0 \cdots 0\ p_0\ p_1\ p_2 \end{vmatrix}_{k-n+1} \quad a_i^{n-2} + U_3 a_i^{n-3} + \cdots + U_{n-1} a_i + U_n = 0,$$

where the U's are polynomials in the p's. If, in the determinant which is the coefficient of a_i^j, we move the elements of the last column to their natural position, the $(k-j)$th column, we obtain $(-1)^{j-n+1}$ multiplying a determinant the form of D^n_{k-n+1} but having for its elements p's instead of V's. If, in the coefficient of a_i^{n-2} we interchange rows and columns and invert the order of the rows and columns, we get a determinant differing from $D^n_{k-n+1}\{j-n+2\}$ only in that we have p's for elements instead of V's. Now apply the corollary to Theorem I to both determinants to change the p's to V's. All of the elements of the coefficient of a_i^j will be positive if we multiply the odd rows and even columns by -1. To accomplish the same thing in the coefficient of a_i^{n-2}, we multiply the first $j-n+2$ rows by -1 and then multiply the odd rows and the even columns by -1. This gives us

$$\begin{aligned} W_0 a_i^k &+ (-1)^{j-n+1}(-1)^{k-n+1} D^n_{k-n+1} a_i^j \\ &+ (-1)^{j-n+2}(-1)^{k-n+1} D^n_{k-n+1}\{j-n+2\} a_i^{n-2} \\ &+ W_3 a_i^{n-3} + \cdots + W_{n-1} a_i + W_n = 0, \end{aligned}$$

where the W's are polynomials in the V's. Solving these equations for the ratio of the two D's, we have

$$\begin{aligned} |k,j,n-3,\cdots,1,0| &: |k,n-2,\cdots,1,0| \\ &= (-1)^{j+k-2n+3} D^n_{k-n+1}\{j-n+2\} : -(-1)^{j+k-2n+2} D^n_{k-n+1}. \end{aligned}$$

Multiplying the left members of the proportion by V_0^{k-n}, we get

$$\begin{aligned} V_0^{k-n}|k,j,n-3,\cdots,1,0| &: V_0^{k-n}|k,n-2,\cdots,1,0| \\ &= D^n_{k-n+1}\{j-n+2\} : D^n_{k-n+1}. \end{aligned}$$

But $V_0^{k-n}|k, n-2, \cdots, 1, 0| = D^n_{k-n+1}$, hence

$$V_0^{k-n}|k,j,n-3,\cdots,1,0| = D^n_{k-n+1}\{j-n+2\}.$$

We may summarize this result in

THEOREM III. *The generalized Vandermonde determinant* $|k, j, n-3, \cdots, 1, 0|$ *multiplied by the* $(k-n)$*th power of its principal is equal to the determinant* $D_{k-n+1}^{n}\{j-n+2\}$ (*see Definition* II).

4. THE GENERAL VANDERMONDIAN

By the use of mathematical induction we shall now prove the formula for the determinant in which we vary any number of exponents of the variables. We note the special case in Theorem III and then proceed to compare $|t_1, t_2, \cdots, t_s, n-s-1, \cdots, 1, 0|$ with $|t_1, t_2, \cdots, t_{s-1}, n-s, \cdots, 1, 0|$. Consider the t_1-n systems of equations

$$\sum_{j=0}^{n} p_j a_i^{l-j} \qquad (l = t_1, t_1-1, \cdots, n\,;\ i = 1,2, \cdots, n).$$

We choose to eliminate all of the variables except $a_i^{t_1}, a_i^{t_2}, \cdots, a_i^{t_{s-1}}, a_i^{t_s}, a_i^{n-s}, a_i^{n-s-1}, \cdots, a_i, 1$, giving us the following (t_1-n+1)th order determinant:

$$\begin{vmatrix}
p_1 & p_2 \cdots p_{t_1-t_2-1} & p_{t_1-t_2+1} \cdots & \cdots & \cdots p_{t_1-t_s-1} & p_{t_1-t_s+1} \cdots \\
p_0 & p_1 \cdots p_{t_1-t_2-2} & p_{t_1-t_2} & \cdots & \cdots p_{t_1-t_s-2} & p_{t_1-t_s} \cdots \\
\cdot & \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot & \cdot & \cdot & \cdot \cdot \cdot \cdot \cdot & \cdot \\
0 & \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot & 0 & p_0 & \cdot \cdot \cdot \cdot & \cdot \\
\cdot & \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot & \cdot & \cdot & \cdot \cdot \cdot \cdot \cdot & \cdot \\
0 & \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot & \cdot & \cdot & 0 & p_0
\end{vmatrix}$$

$$\begin{vmatrix}
p_n\ 0 \cdots 0 & p_0 a_i^{t_1} + p_{t_1-t_2} a_i^{t_2} + \cdots + p_{t_1-t_s} a_i^{t_s} \\
\cdot\ p_n\ 0 \cdots 0 & p_{t_1-t_2-1} a_i^{t_2} + \cdots + p_{t_1-t_s-1} a_i^{t_s} \\
\cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot & \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \\
\cdot \cdot \cdot\ p_{n-1} & \cdots + p_{2n-s-t_s} a_i^{t_s} + p_n a_i^{n-s} \\
\cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot & \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \\
\cdot \cdot \cdot\ p_{s-1} & p_{n-t_s} a_i^{t_s} + p_s a_i^{n-s} + \cdots
\end{vmatrix}_{t_1+n+1} = 0,$$

where it is to be understood that $p_k = 0$, whenever $k > n$. Writing this determinant as the sum of $n+1$ determinants by regarding the last column as the sum of n columns each of which has the same power of a in all of its elements, we obtain

$$M_1 a_i^{t_1} + M_2 a_i^{t_2} + \cdots + M_{s-1} a_i^{t_{s-1}}$$

$$+ \begin{vmatrix}
p_1 & p_2 \cdots & \cdots & \cdots p_{t_1-t_s-1} & p_{t_1-t_s+1} \cdots p_n & 0 \cdots 0 & p_{t_1-t_s} \\
p_0 & p_1 \cdots & \cdots & \cdots p_{t_1-t_s-2} & p_{t_1-t_s} \cdots p_n & 0 \cdots 0 & p_{t_1-t_s-1} \\
\cdot & \cdot \cdot \cdot \cdot & \cdot \cdot \cdot & \cdot \cdot \cdot \cdot \cdot \cdot \cdot & \cdot \cdot \cdot \cdot \cdot \cdot \cdot & \cdot \cdot \cdot \cdot \cdot \cdot \cdot & \cdot \\
0 & \cdots 0\ p_0 & \cdots & p_{n-t_s-1}\ p_{n-t_s+1} & \cdots \cdots & p_{s-1}\ p_{n-t_s} &
\end{vmatrix}_{t_1-n+1} a_i^{t_s}$$

$$
\begin{vmatrix}
p_1 & p_2 & \cdots & \cdots & \cdots & p_{t_1-t_s-1} & p_{t_1-t_s+1} & \cdots & p_n & 0 & \cdots & 0 & 0 \\
p_0 & p_1 & \cdots & \cdots & \cdots & p_{t_1-t_s-2} & p_{t_1-t_s} & \cdots & p_n & 0 & \cdots & 0 & 0 \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
0 & \cdots & \cdots & \cdots & \cdots & 0 & p_0 & \cdots & \cdots & \cdots & p_{n-1} & p_n & a_i^{n-s} \\
0 & \cdots & \cdots & \cdots & \cdots & 0 & p_0 & \cdots & \cdots & \cdots & p_{n-2} & p_{n-1} & \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \\
0 & \cdots & \cdots & \cdots & \cdots & & & & 0 & p_0 & \cdots & p_{s-2} & p_s \quad t_1-n+1
\end{vmatrix}
$$

$$
+ N_{s+1}a_i^{n-s-1} + N_{s+2}a_i^{n-s-2} + \cdots + N_{n-1}a_i + N_n = 0,
$$

where the M's and the N's are polynomials in the p's. Let us consider the determinants which are the coefficients of $a_i^{t_s}$ and a_i^{n-s}. Place the elements of the last column of the coefficient of $a_i^{t_s}$ in their natural position, the (t_1-t_s-s+2)th column; then interchange rows and columns; then invert the order of the rows and columns, giving us $(-1)^{t_s+s-n-1}$ multiplying a determinant of the form of $D_{t_1-n+1}\{t_2-n+2, t_3-n+3, \cdots, t_{s-1}-n+s-1\}$, but having for its elements p's in place of V's. If, in the coefficient of a_i^{n-s}, we interchange rows and columns and then invert the order of rows and columns, we get a determinant differing from $D^n_{t_1-n+1}\{t_2-n+2, t_3-n+3, \cdots, t_s-n+s\}$ only in that the elements are p's instead of V's. We now apply the corollary to Theorem I to both determinants to change the p's to V's. All of the elements of the coefficient of $a_i^{t_s}$ will be positive if we multiply the first t_2-n+2 rows by -1, the first t_3-n+3 rows again by -1, $\cdots$ the first $t_{s-1}-n+s-1$ rows again by -1; and then multiply the odd rows and even columns by -1. This requires $(\sum_{j=2}^{s-1} t_j-n+j)+t_1-n+1=\sum_{j=1}^{s-1}t_j-n+j$ changes in sign. When this number is added to the $t_s-n+s-1$ changes mentioned above, we get $-1+\sum_{j=1}^{s}t_j-n+j$ changes. We can make all of the changes of the coefficient of a_i^{n-s} positive by multiplying the first t_2-n+2 rows by -1, the first t_3-n+3 rows again by -1, $\cdots$ the first t_s-n+s rows again by -1; and then multiplying the odd rows and the even columns by -1. This requires $\sigma\equiv\sum_{j=1}^{s}t_j-n+j$ changes in sign. Our equation then takes the form

$$
\begin{aligned}
&X_1a_i^{t_1} + X_2a_i^{t_2} + \cdots + X_{s-1}a_i^{t_{s-1}} \\
&\quad + (-1)^{-1+\sigma}D^n_{t_1-n+1}\{t_2+n+2, t_3+n+3, \cdots, t_{s-1}-n+s-1\}a_i^{t_s} \\
&\quad + (-1)^{\sigma}D^n_{t_1-n+1}\{t_2-n+2, t_3-n+3, \cdots, t_s-n+s\}a_i^{n-s} \\
&\quad + Y_{s+1}a_i^{n-s-1} + Y_{s+2}a_i^{n-s-2} + \cdots + Y_{n-1}a_i + Y_n = 0,
\end{aligned}
$$

where the X's and the Y's are polynomials in the V's. We can then say

$$\begin{aligned}|\, t_1,t_2,\cdots,t_s,n-s-1,\cdots,1,0\,| : |\, t_1,t_2,\cdots,t_{s-1},n-s,\cdots,1,0\,| \\ = D^n_{t_1-n+1}\{t_2-n-2,t_3-n+3,\cdots,t_s-n+s\} \\ : D^n_{t_1-n+1}\{t_2-n+2,t_3-n+3,\cdots,t_{s-1}-n+s-1\}.\end{aligned}$$

Multiplying the left members of the proportion by $V_0{}^{t_1-n}$ and noting the special case where $s=2$, and $t_{s-1}=k$, and $t_s=j$ in the preceding theorem, we have

$$\begin{aligned}V_0{}^{t_1-n}|\, t_1,t_2,\cdots,t_s,n-s-1,\cdots,1,0\,| \\ = D^n_{t_1-n+1}\{t_2-n+2,t_3-n+3,\cdots,t_s-n+s\}.\end{aligned}$$

We may now state our general theorem:

THEOREM IV. *The generalized Vandermonde determinant* $|t_1, t_2, \cdots, t_s, n-s-1, \cdots, 1, 0|$ *multiplied by the* (t_1-n)*th power of the principal Vandermondian is equal to the determinant* $D^n_{t_1-n+1}\{t_2-n+2, t_3-n+3, \cdots, t_s-n+s\}$.

5. SYMMETRIC FUNCTIONS

We shall now use the preceding results to devise a method for expressing any integral rational symmetric function in terms of elementary symmetric functions. To accomplish this, we make use of the following theorem of Muir's: The product of a simple alternant and a single symmetric function of its variables is expressible by a sum of simple alternants, whose indices are got by arranging the variables in every term of the symmetric function in the same order, and adding the indices of each term to the indices of the original alternant, the first to the first, the second to the second, and so on.* For example

$$\begin{aligned}|\, a^0b^1c^2\,|\ \textstyle\sum a^3b = |\, a^3b^2c^2\,| + |\, a^3b^1c^3\,| + |\, a^1b^4c^2\,| + |\, a^0b^4c^3\,| \\ + |\, a^1b^1c^5\,| + |\, a^0b^2c^5\,| = |\, a^0b^2c^5\,| - |\, a^0b^3c^4\,| - |\, a^1b^2c^4\,|.\end{aligned}$$

Using this theorem, we can express as a function of generalized Vandermondians the product of any integral rational symmetric function by the principal Vandermondian in the same number or larger number of variables. Since every generalized Vandermondian is expressible in terms of principal and secondary Vandermondians, and these in turn are expressible in terms of elementary symmetric functions, we can express every integral rational symmetric function in terms of elementary symmetric functions. For

* Muir's *Theory of Determinants*, vol. IV, p. 151.

example, to compute $\sum a^2bc$ in four variables, we multiply the principal Vandermondian in four variables by the given function:

$$|a^0b^1c^2d^3| \cdot \sum a^2bc = |a^2b^1c^3d^4| + |a^1b^3c^2d^4| + |a^1b^2c^4d^3| + |a^0b^2c^3d^5| = |a^0b^2c^3d^5| - 3|a^1b^2c^3d^4|.$$

Using the notation used in the preceding sections,

$$V_0 \sum a^2bc = |5\,3\,0\,2| - 3|4\,3\,2\,1| = \frac{\begin{vmatrix} V_3 & V_4 \\ V_0 & V_1 \end{vmatrix}}{V_0} - 3V_4$$

or

$$\sum a^2bc = \frac{V_1V_3 - V_0V_4}{V_0^2} - \frac{3V_4}{V_0} = E_1E_3 - 4E_4.$$

We shall now prove that the result for n variables, where $n>r$, can be obtained if a symmetric function of degree r is multiplied by the principal Vandermondian in r variables. That is to say, when we express $\sum a_0^{m_0}a_1^{m_1} \cdots a_i^{m_t}$ for n variables, where $m_0 \geqq m_1 \geqq \cdots \geqq m_t$ and $\sum_{i=0}^{t} m_i = r$, in terms of elementary symmetric functions, none of the subscripts of the E's can exceed r. To show this, we multiply $|a_0^0a_1^1 \cdots a_{n-1}^{n-1}|$ by $\sum a_0^{m_0}a_1^{m_1} \cdots a_i^{m_t}$. The product will consist of a number of generalized Vandermondians. The highest exponent which can occur in the symbols for any of these general Vandermondians is m_0+n-1 and the number of exponents in these symbols which are increased above what they are for the principal Vandermondian is at most t, since there are only t exponents in the given symmetric function. Hence the highest subscript in any of the D's is $m_0+n-1-n+1$ and no D followed by a brace can have more than t elements in the brace. Hence the highest subscript in any V can not exceed m_0+t, but since $m_1, m_2, \cdots, m_t$ are all greater than or equal to 1, it follows that $r = m_0 + \sum_{i=1}^{t} m_i \geqq m_0 + t$, so that no subscript of any V can exceed r.

Before concluding the paper, it may not be amiss to present another example. In order to express $\sum a^3b^2$ in terms of elementary symmetric functions, we multiply it by the principal Vandermondian in 5 variables:

$$|a^0b^1c^2d^3e^4| \sum a^3b^2 = |a^3b^1c^2d^5e^4| + |a^0b^4c^2d^3e^6| + |a^2b^1c^5d^3e^4| + |a^0b^1c^5d^3e^6| + |a^0b^3c^2d^6e^4| + |a^0b^1c^4d^3e^7| + |a^0b^1c^2d^5e^7|.$$

Changing notation, we get

$$V_0 \sum a^3b^2 = |75210| - |74310| - |65310| + 2|64320| - 2|54321|$$

$$= \frac{\begin{vmatrix} V_2 & V_3 & V_4 \\ V_1 & V_2 & V_3 \\ 0 & V_0 & V_1 \end{vmatrix}}{V_0^2} - \frac{\begin{vmatrix} V_3 & V_4 & V_5 \\ V_0 & V_1 & V_2 \\ 0 & V_0 & V_1 \end{vmatrix}}{V_0^2} - \frac{\begin{vmatrix} V_3 & V_4 \\ V_1 & V_2 \end{vmatrix}}{V_0} + 2\frac{\begin{vmatrix} V_4 & V_5 \\ V_0 & V_1 \end{vmatrix}}{V_0} - 2V_5$$

$$= \frac{V_1V_2^2 - 2V_1^2V_3 - V_0V_2V_3 + 5V_0V_1V_4 - 5V_0^2V_5}{V_0^2}.$$

Or

$$\sum a^3b^2 = E_1E_2^2 - 2E_1^2E_3 - E_2E_3 + 5E_1E_4 - 5E_5.$$

TEXAS TECHNOLOGICAL COLLEGE,
LUBBOCK, TEXAS

ON APPROXIMATION BY RATIONAL FUNCTIONS TO AN ARBITRARY FUNCTION OF A COMPLEX VARIABLE*

BY

J. L. WALSH

1. **Introduction.** In his classical paper† on approximation to analytic functions of a complex variable by means of polynomials and other rational functions, Runge proved two important theorems:

THEOREM I. *If the function $f(z)$ is analytic in a closed Jordan region of the z-plane, then in that closed region $f(z)$ can be approximated as closely as desired by a polynomial in z.*

THEOREM II. *If the function $f(z)$ is analytic in a closed region R of the z-plane which is bounded by a finite number n of Jordan curves of which no two have a common point, then in that closed region $f(z)$ can be approximated as closely as desired by a rational function of z.*

Cauchy's Integral for the function and region (if the Jordan curves bounding R are rectifiable) in Theorem II splits up the function $f(z)$ into n functions. If Theorem I is applied to each of these in turn (a simple transformation $w=1/(z-\alpha_i)$ is first necessary for $n-1$ of them), we find Theorem II immediately.

If the Jordan curves bounding R are not rectifiable, we readily obtain the same result. Consider an auxiliary variable region R' of the same connectivity as R, which lies within R, which is bounded by n rectifiable Jordan curves, and which approaches R as the bounding curves change continuously. For the region R', Cauchy's Integral for $f(z)$ splits that function into n functions *which do not change as R' varies*. Each of these n functions is, when properly defined, analytic not merely in R but in a region bounded by a single Jordan curve and which contains R. Theorem II then follows as before from Theorem I.

On the possibility of approximation of arbitrary functions by polynomials, no deeper lying results than Theorems I and II, except for approximation by certain *restricted* types of polynomials, appear to have been found until

---

* Presented to the Society, December 27, 1928; received by the editors October 12, 1928, and February 8, 1929.

† Acta Mathematica, vol. 6 (1885), pp. 229–244.

recently, when the problem was attacked by use of the modern theory of conformal mapping. We mention generalizations of Theorems I and II:*

THEOREM III. *If the function $f(z)$ is analytic interior to a Jordan curve and continuous in the corresponding closed region, then in that closed region $f(z)$ can be approximated as closely as desired by a polynomial in z.*

THEOREM IV. *If the function $f(z)$ is analytic interior to a region which is bounded by a finite number of Jordan curves of which no two have a common point, and is continuous in the corresponding closed region, then in that closed region $f(z)$ can be approximated as closely as desired by a rational function of z.*

These theorems refer to functions which are analytic in some region, but there are similar results which hold for functions known merely to be continuous:†

THEOREM V. *If the function $f(z)$ is continuous on a Jordan arc, then on that arc $f(z)$ can be approximated as closely as desired by a polynomial in z.*

THEOREM VI. *If the function $f(z)$ is continuous on a Jordan curve in whose interior the origin lies, then on that curve $f(z)$ can be approximated as closely as desired by a polynomial in z and $1/z$.*

Theorem V is a direct generalization of Weierstrass's classical theorem on the possibility of approximation by a polynomial to a real continuous function in a given interval of the axis of reals, and Theorem VI is a generalization of Weierstrass's theorem on the possibility of approximation to a real continuous function in an interval of the axis of reals by trigonometric polynomials.‡ For if the Jordan curve of Theorem VI is the unit circle, we have

$$z^n = \cos n\theta + i \sin n\theta, \qquad z^{-n} = \cos n\theta - i \sin n\theta,$$

$$\sin n\theta = \frac{z^n - z^{-n}}{2i}, \qquad \cos n\theta = \frac{z^n + z^{-n}}{2},$$

so that every trigonometric polynomial is a polynomial in z and $1/z$, and conversely.

2. **Principal results to be obtained.** In the present paper we shall establish a result which is of great generality and includes all the results

* Walsh, Mathematische Annalen, vol. 96 (1926), pp. 430–436.

† Walsh, Mathematische Annalen, vol. 96 (1926), pp. 437–450.

‡ To be sure, Weierstrass considers merely real functions and real approximations, but there is no difficulty in proceeding from a complex approximation to a real function to a real approximation to that function and vice versa.

just mentioned. The proof is based primarily on results already attained, and uses in addition only a few elementary transformations of the plane.

We notice that uniform approximation with an arbitrarily small error is equivalent to uniform expansion in series, and frequently the latter language is more convenient for us. As a matter of convention, the point at infinity is to be considered adjoined to the plane whenever we are dealing with approximation by general rational functions (i.e., not by polynomials), but is not to be so considered when we are dealing with approximation by polynomials, even though these polynomials may be, for instance, polynomials in z and $1/z$. When we are dealing with rational functions, a Jordan region is thus an arbitrary region *on the sphere* (stereographic projection of the plane) bounded by a simple closed curve. Likewise as a matter of notation, if a region or other open point set is denoted by such a symbol as C or R, then the corresponding closed region or closed point set is denoted by $\bar{C}$ or $\bar{R}$.*

Suppose that on a given closed point set it is desired to approximate by rational functions a given function $f(z)$ analytic on that closed point set. This point set may be so closely approximated by a region bounded entirely by Jordan curves, of which no two have a common point, that $f(z)$ is analytic in this new closed region, and so that the given point set is contained in this new region. Theorem II yields directly

THEOREM VII. *An arbitrary function $f(z)$ analytic on a closed point set can be uniformly approximated on that point set as closely as desired, by a rational function of z.*

If we do not assume analyticity of the function on the closed point set, we shall have to content ourselves with a less general point set:

THEOREM VIII. *Let M be a closed point set whose boundary consists of a finite number of Jordan arcs which do not separate the plane into more than a finite number of regions. Then an arbitrary function $f(z)$ continuous on M and analytic in the interior points of M can be uniformly approximated on M as closely as desired by a rational function of z.*

If M divides the plane into precisely k regions none of which belongs to M, this rational function can be chosen so that its only poles lie in k preassigned points, one in each of these regions.

This is the most interesting result of the present paper; most of the remainder of the paper is devoted to its proof. Theorem VIII contains Theorems I-VI as special cases. Another special case worth mentioning is

* For typographical reasons, if the given region is C' or C_1, the corresponding closed region will be denoted by $\bar{C}'$ or $\bar{C}_1$, the bar not extending over the subscript or superscript.

THEOREM IX. *If M is a closed point set composed of a finite number of Jordan arcs which divide the plane into at most a finite number of regions, then an arbitrary function continuous on M can be uniformly approximated on M as closely as desired by a rational function of z.*

We state merely for the sake of reference two other results on approximation which have been previously proved and are to be used in the sequel:

THEOREM X. *If on the closed point set M the function $f(z)$ can be uniformly approximated as closely as desired by a rational function, and if there be chosen n arbitrary points of M, then there exists a rational function which on M approximates $f(z)$ as closely as desired and which coincides with $f(z)$ in these n points.**

THEOREM XI. *If the point set M consists of the closed interiors of two Jordan regions with but a single common point, then an arbitrary function $f(z)$ analytic in the interior points of M and continuous on M can be uniformly approximated on M as closely as desired by a polynomial in z.*†

We proceed now to prove a sequence of theorems leading up to Theorem VIII.

3. **Regions bounded by two Jordan curves.** We shall prove the following theorem:

THEOREM XII. *Let the Jordan curve C_2 contain the origin in its interior and itself lie interior to the Jordan curve C_1, except that C_1 and C_2 have the single point $z=\alpha$ in common. Denote by R the region bounded by C_1 and C_2. Then an arbitrary function $f(z)$ analytic in R and continuous in $\overline{R}$ can be developed in $\overline{R}$ uniformly in a series of polynomials in z and $1/z$.*

We shall prove the theorem first in a particular case, but the general case can be reduced to this particular case by a number of intermediate steps which we treat in detail.

Case I. *The curves C_1 and C_2 have continuously turning tangents and continuous curvature, and the function $f(z)$ satisfies a Lipschitz condition*‡ *on C_1 and C_2.* The function $f(z)$ can be represented in R by Cauchy's Integral

---

* Walsh, these Transactions, vol. 30 (1928), pp. 307–332; Theorem X. The theorem is there proved for the case of approximation by polynomials, but the proof holds, after a suitable transformation $w=1/(z-\alpha)$ if M contains the point at infinity, also in the present case, even if $f(z)$ becomes infinite on M. We notice, as an essential part of the proof, that the theorem remains true if M is the entire plane, for in that case $f(z)$ must itself be a rational function.

† Walsh, these Transactions, vol. 30 (1928), pp. 472–482; Theorem III.

‡ Here may be taken in the form $|f(z_1)-f(z_2)| \leqq M|z_1-z_2|$, M fixed, z_1 and z_2 on C_1 or on C_2.

$$f(z) = f_1(z) - f_2(z), \tag{1}$$

$$f_1(z) = \frac{1}{2\pi i}\int_{C_1} \frac{f(t)dt}{t - z}, \qquad f_2(z) = \frac{1}{2\pi i}\int_{C_2} \frac{f(t)dt}{t - z},$$

where both integrals are taken in the positive (i.e., counterclockwise) sense on the curves. But with the present restrictions on the curves and on the function $f(z)$, it follows from a theorem due to Plemelj* that the function $f_1(z)$ just defined is analytic interior to C_1 and continuous in the corresponding closed region (when properly defined on C_1) and that the function $f_2(z)$ is analytic exterior to C_2 and continuous in the corresponding closed region, even at infinity, when it is properly defined on C_2 and at infinity. By virtue of Theorem III† the function $f_1(z)$ can be expanded uniformly in the closed interior of C_1 by polynomials in z, and the function $f_2(z)$ can be expanded uniformly in the closed exterior of C_2 by polynomials in $1/z$. Equation (1) holds *on* the curves C_1 and C_2, for the functions $f(z)$, $f_1(z)$, $f_2(z)$ are all continuous in $\overline{R}$, so our theorem is completely established in Case I.

In this particular case the given function $f(z)$ can be expressed in $\overline{R}$ as the sum of two functions, which in $\overline{R}$ can be respectively represented as a series of polynomials in z and as a series of polynomials in $1/z$. This simple result is not to be expected in the more general case.

Case II. *The curves C_1 and C_2 are tangent at $z=\alpha$ and circles can be drawn tangent to them at $z=\alpha$ which contain respectively no point of the interior of C_1 and no point of the exterior of C_2.* Here there can be drawn a Jordan curve C_1' with a continuously turning tangent and continuous curvature which is tangent to C_1 at $z=\alpha$ and contains all other points of C_1 in its interior. There can likewise be drawn a Jordan curve C_2' which has a continuously turning tangent and continuous curvature which is tangent to C_2 at $z=\alpha$, whose exterior contains all other points of C_2, and which contains the origin in its interior.‡ We shall show how there may be defined a function $\phi(z)$ which differs from $f(z)$ in $\overline{R}$ by an amount less than any preassigned positive quantity ϵ and which is analytic in the region R' bounded by C_1' and C_2', contin-

---

* Monatshefte für Mathematik und Physik, vol. 19 (1908), pp. 205–210.

† In connection with $f_2(z)$ we need to make a reciprocal transformation $w=1/z$ or to apply the theorem given by the present writer, Mathematische Annalen, vol. 96 (1926), p. 435.

‡ The curves C_1' and C_2' are readily drawn, say by using arcs of the two circles mentioned in the hypothesis on C_1 and C_2, by using broken lines in conjunction with those arcs to construct the curve desired, finally by smoothing out the corners. The latter process is readily accomplished, so far as concerns the intersection of two straight lines, by the use of arches of the curve $y=a \sin bx$, $0 \leqq x \leqq \pi/b$, where a and b are suitably chosen. The same curve (but not an entire arch) can also be used to smooth out the intersection of a straight line and a circle.

uous in the corresponding closed region, and which satisfies a Lipschitz condition on the curves C_1' and C_2'. Transform the plane by means of the transformation $w=z^{1/2}$, so that the regions R and R' are transformed into regions S and S' respectively. In reality each of the regions R and R' is of course transformed into *two* regions in the w-plane, but we fasten our attention merely on one pair of such regions, resulting from transformation by a single branch of the function w. It follows from Theorem III that when $\epsilon>0$ is given, there exists a function $\Phi(w)$ analytic in $\bar{S}'$ and which differs in $\bar{S}$ from $f(w^2)$ by less than ϵ. This function can be chosen, by Theorem X, so that the two values $\Phi(+\alpha^{1/2})$ and $\Phi(-\alpha^{1/2})$ are equal, in fact both equal to $f(a)$. We transform back to the z-plane by the transformation $z=w^2$, and consider the z-plane cut along a curve from the origin to the point $z=\alpha$, the curve having no point in common with (the open region) R'. The function $\Phi(w)$ is transformed into a function $\phi(z)=\Phi(z^{1/2})$, where $z^{1/2}$ is a properly determined branch of that function, and this function $\phi(z)$ is single-valued in $\bar{R}'$ and has in $\bar{R}'$ the properties mentioned in the requirements for Case I. We notice that $\phi(z)$ is analytic in every point of $\bar{R}'$ except $z=\alpha$ and continuous in that closed region, but we omit the details of the proof that $\phi(z)$ satisfies a Lipschitz condition on C_1' and on C_2'. We have Case I, however, and the function $\phi(z)$ can be uniformly approximated in $\bar{R}'$ and hence in $\bar{R}$ as closely as desired by means of a polynomial in z and $1/z$; therefore the function $f(z)$, which differs from $\phi(z)$ in $\bar{R}$ by less than ϵ, can in $\bar{R}$ be uniformly approximated as closely as desired by such a polynomial, and Case II of the theorem is established.

Case III. *The general case.* By a series of transformations we proceed to bring the general case under Case II. We assume, as we may do with no loss of generality, that $|z-\alpha|<1$ for all z in $\bar{R}$. The first transformation to be used is

$$z-\alpha=e^{1/(w-\alpha)}$$

which transforms R into an infinite number of regions in the w-plane, of which we consider a single region R_1, necessarily of like connectivity to R. That is, R_1 is bounded by two Jordan curves with the single point $w=\alpha$ in common. The effect of this transformation is perhaps best seen by breaking it up into the two transformations $z-\alpha=e^{z'}$, $z'=1/(w-\alpha)$. The former transformation carries $z=\alpha$ into $z'=\infty$, but any transform of R onto the z'-plane lies in the half-plane to the left of the axis of imaginaries, for $|z-\alpha|<1$ in R. The transformation $z'=1/(w-\alpha)$ carries the axis of imaginaries in the z'-plane into a straight line through $w=\alpha$, so the new transform of R (i.e. in the w-plane) is a region R_1 of like connectivity to R, as stated. More-

over the region R_1 is finite and lies in a half-plane bounded by a line through $w=\alpha$.

We now use again this same transformation, setting

$$w - \alpha = e^{1/(u-\alpha)}.$$

Under the first part of this transformation, $w-\alpha=e^{w'}$, the region R_1 is transformed into a region Σ of the w' plane (we consider only one of the infinitely many transforms of R_1) which lies not merely in a half-plane bounded by a parallel to the axis of imaginaries, but also in a strip of width π bounded by two parallels to the axis of reals. Then the transformation $w'=1/(u-\alpha)$ carries the region Σ into a region R_2 which lies between two circles which are tangent at the point $u=\alpha$. The region R_2 is of course finite, for $u=\infty$ corresponds to $w-\alpha=1$, $z-\alpha=e$, which is impossible in $\overline{R}$. The region R_2 lies, in the neighborhood of $u=\alpha$, entirely on one side of the common perpendicular to these circles at $u=\alpha$. Suppose R_2 to be bounded by the Jordan curves Γ_1 and Γ_2, with Γ_2 interior, except for the single common point $u=\alpha$, to the curve Γ_1.

We transform now to the t-plane by setting

$$\frac{t-\alpha}{t} = \left(\frac{u-\alpha}{u-\beta}\right)^{1/2},$$

where β is an arbitrarily chosen point interior to Γ_2. If the u-plane is considered cut along a Jordan arc joining α and β and which contains no point of $\overline{R}_2$ other than α, the region R_2 is transformed into two regions of the t-plane, of which we consider a single finite region T. The region T is bounded by two Jordan curves C_1' and C_2', so that C_2' is interior to C_1' except for the single point $t=\alpha$ which the two curves have in common. Moreover a circle can be drawn tangent to C_1' and C_2' at $t=\alpha$ which contains no point of the interior of C_1', and a circle can be drawn tangent to C_1' and C_2' at $t=\alpha$ which contains no point of the exterior of C_2'. In fact the transformation just used is, except for linear transformation of the two planes which transforms circles into circles, of the form

$$y = x^{1/2}.$$

A circle in the x-plane tangent to the axis of reals at the origin is transformed into a curve* one of whose branches in the neighborhood of the origin lies between that circle and the axis of reals. For under the transformation

* A lemniscate as a matter of fact, but for our present purpose we do not need such detailed information.

$y=x^{1/2}$, the argument of the complex quantity x is halved, the modulus raised to the power one-half.

The transformation $x=(u-\alpha)/(u-\beta)$ is here used to transform R_2 into a region R_2' of slightly different type. If a point tracing either Γ_1 or Γ_2 leaves the curve at $u=\alpha$ and traces the tangent without reversing the sense of motion, the point moves *exterior* to Γ_1 and Γ_2. If a point traces either bounding curve of R_2', leaves the curve at $x=0$ and traces the tangent without reversing the sense of motion, the point moves *interior* to those curves. Such a transformation as the one used is necessary to ensure that R_2' is the latter of these two types of regions instead of the former, giving us the proper region T.

The region T satisfies the conditions imposed in Case II, except that we do not know that the origin $t=0$ lies interior to C_2'. Hence the transform of $f(z)$, which is analytic in T and continuous in $\overline{T}$, can be uniformly developed in $\overline{T}$ in a series of polynomials in t and $1/(t-\gamma)$, where γ is a suitably chosen point interior to C_2'. But in the u-plane, the functions t and $1/(t-\gamma)$ (i.e., suitably chosen branches) are analytic in a region bounded by Γ_2 and an arbitrary Jordan curve containing Γ_1, and are continuous in the corresponding closed region. Hence in that closed region, which contains $\overline{R}_2$, the functions t and $1/(t-\gamma)$ can each be developed into a uniformly convergent series of polynomials in u and $1/(u-\beta)$. Thus the transform of $f(z)$ can be developed uniformly in such a series of polynomials.

In our transformation from the z-plane to the u-plane it is convenient to think of the z-plane as cut along a Jordan curve from the point $z=\alpha$ to infinity; we suppose that this curve has no point other than $z=\alpha$ in common with $\overline{R}$. Then when u is properly chosen as a branch of the function of z defined by

$$z-\alpha=e^{1/(w-\alpha)}, \qquad w-\alpha=e^{1/(u-\alpha)},$$

the functions u and $1/(u-\beta)$ are analytic in an annular region which is bounded by C_1 and by a second Jordan curve interior to C_2 and are continuous in the corresponding closed region. This annular region is thus chosen so that it contains R, but is of course so chosen that it does not contain the transform of $u=\beta$. In the closed annular region the functions u and $1/(u-\beta)$ can be uniformly expanded in series of polynomials in z and $1/z$; this expansion is valid likewise in $\overline{R}$, so $f(z)$ itself can in $\overline{R}$ be uniformly expanded in a series of polynomials in z and $1/z$, and the theorem is completely established.

4. **Regions bounded by several Jordan curves.** Theorem XII is to be used in proving a more general result:

THEOREM XIII. *If the region R is bounded by a finite number of Jordan curves, then an arbitrary function $f(z)$ analytic in R and continuous in $\overline{R}$ can be uniformly expanded in $\overline{R}$ in a series of rational functions of z.*

Our proof here is by induction. We suppose the theorem true for a region bounded by n Jordan curves and shall prove the theorem true for a region bounded by $n+1$ such curves. Any region bounded by two Jordan curves is included either under Theorem IV or Theorem XII, so the induction is started.

Suppose R to be the region interior to a Jordan curve C_0 and exterior to the Jordan curves $C_1, C_2, \cdots, C_n$. If none of these latter curves has a point of intersection with C_0, we proceed as in Theorem II. Cauchy's Integral enables us to express the given function $f(z)$ as the sum of two functions, the one analytic interior to C_0 and continuous in the corresponding closed region, the other function analytic exterior to $C_1, C_2, \cdots, C_n$ and continuous in the corresponding closed region. Each of these functions can be expanded in a series of rational functions uniformly convergent in the closed regions indicated, so the same is true of $f(z)$ in $\overline{R}$.

We assume, then, as we may do with no loss of generality, that say C_1 has a point $z=\alpha$ in common with C_0. Let β be a point interior to C_1, consider the z-plane cut from β to infinity along a Jordan arc which has no point other than α in common with $\overline{R}$, and transform $\overline{R}$ by means of the transformation $z-\beta=w^2$. A properly chosen branch of the function w transforms $\overline{R}$ into a region $\overline{R}'$ and $f(z)$ into a function $\phi(w)$ analytic interior to R' and continuous in $\overline{R}'$. The region R' is bounded by the transforms $C_2', C_3', \cdots, C_n'$ of the curves $C_2, C_3, \cdots, C_n$, and by the Jordan curve $C_0'+C_1'$, the transform of the curves C_0 and C_1. If an arbitrary $\epsilon>0$ be given, then by our theorem as assumed for a region bounded by n Jordan curves, there exists a rational function $\Phi(w)$ which is analytic in $\overline{R}'$, which in $\overline{R}'$ differs from $\phi(w)$ by a function in absolute value not greater than ϵ, and such that $\Phi(+(\alpha-\beta)^{1/2}) = \Phi(-(\alpha-\beta)^{1/2})$; this last fact results from Theorem X.

The function $\Phi(w)$ is therefore analytic in a region $\overline{S}'$ bounded by $C_0'+C_1'$ and by certain Jordan curves $\Gamma_2', \Gamma_3', \cdots, \Gamma_n'$ which are *interior* respectively to $C_2', C_3', \cdots, C_n'$. We transform now back to the z-plane; the function $\Phi(w)$ goes into a function $F(z)$ which differs from $f(z)$ in $\overline{R}$ by at most ϵ, and this function $F(z)$ is analytic interior to a region (which includes R) bounded by C_0, C_1, and by the transforms $\Gamma_2, \Gamma_3, \cdots, \Gamma_n$ of $\Gamma_2', \Gamma_3', \cdots, \Gamma_n'$ respectively. The function $F(z)$ is continuous in the corresponding closed region. Cauchy's Integral enables us as in Theorem II to express $F(z)$ in $\overline{R}$ as the sum of $n-1$ functions analytic respectively in

the closed regions exterior to $\Gamma_2, \Gamma_3, \cdots, \Gamma_n$ and analytic in the region bounded by C_0 and C_1, continuous in the corresponding closed region. Each of these functions forming the sum is expressible in the closed region corresponding to it as a uniformly convergent series of rational functions of z. Hence $F(z)$ and $f(z)$ are so expressible in $\overline{R}$.

5. **Two regions with common points.** Somewhat similar to Theorem XIII is

THEOREM XIV. *Let the point set R be composed of the interiors of two Jordan curves C_1 and C_2 which lie exterior to each other except for n points which they have in common. Then an arbitrary function $f(z)$ analytic in R and continuous in $\overline{R}$ is uniformly developable in $\overline{R}$ in a series of rational functions of z.*

The case $n=1$ is essentially the case of Theorem XI. The proof for the general case is by induction, assuming the result true for n common points and proving the result for $n+1$ common points. The proof is moreover so similar to the proof of Theorem XIII that it will be merely sketched. The point set R may be considered interior to the bounding Jordan curve D_0 and exterior to the bounding Jordan curves $D_1, D_2, \cdots, D_n$; each of these curves is composed of an arc of C_1 and an arc of C_2. Assume D_1 to have a point $z=\alpha$ in common with D_0 and choose a point β interior to D_1. Cut the z-plane as before and transform by setting $z-\beta=w^2$. The present theorem for n common points yields a function differing only slightly from $f(z)$, analytic in the transform of the point set R, taking on the same values in the two points $+(\alpha-\beta)^{1/2}$ and $-(\alpha-\beta)^{1/2}$. Transformation back to the z-plane gives a function differing in $\overline{R}$ only slightly from $f(z)$, and analytic in $\overline{R}$ except possibly at α. This new function is analytic interior to a regon bounded by D_0 and D_1, and by curves $\Delta_2, \Delta_3, \cdots, \Delta_n$ respectively interior to $D_2, D_3, \cdots, D_n$, and is continuous in the corresponding closed region. This function is therefore the sum of n functions, analytic in the regions bounded by D_0 and D_1, and by $\Delta_2, \Delta_3, \cdots, \Delta_n$ respectively, and continuous in the corresponding closed regions. Each of these n functions can be expanded in $\overline{R}$ in a series of rational functions of z, so that is true also of $f(z)$.

6. **Approximation in two regions with common points.** Theorem XIV leads almost at once to

THEOREM XV. *Let the two Jordan curves C_1 and C_2 lie exterior to each other except for the k points $\alpha_1, \alpha_2, \cdots, \alpha_k$ which they have in common. Let functions $f_1(z)$ and $f_2(z)$ defined in $\bar{C}_1$ and $\bar{C}_2$ respectively or more generally on point sets contained in $\bar{C}_1$ and $\bar{C}_2$ be uniformly developable in $\bar{C}_1$ and $\bar{C}_2$ re-*

spectively (or on those point sets) in rational functions of z. Let moreover both $f_1(\alpha_i)$ *and* $f_2(\alpha_i)$ *be defined and let the equations*

$$f_1(\alpha_i) = f_2(\alpha_i) \qquad (i = 1, 2, \cdots, k)$$

hold. Then there exists a series of rational functions of z converging to the sum $f_1(z)$ *uniformly in* $\bar{C}_1$ *(or on the sub-set) and to the sum* $f_2(z)$ *uniformly in* $\bar{C}_2$ *(or on the sub-set).*

We write the given sequences in the form

$f_1(z) = \lim_{n\to\infty} (f_n'(z) + \phi_n'(z))$, uniformly on $\bar{C}_1$ or sub-set,

$f_2(z) = \lim_{n\to\infty} (f_n''(z) + \phi_n''(z))$, uniformly on $\bar{C}_2$ or sub-set,

where the poles of $f_n'(z)$ and $f_n''(z)$ lie in $\bar{C}_1$ and $\bar{C}_2$ respectively, and the poles of $\phi_n'(z)$ and $\phi_n''(z)$ lie outside of $\bar{C}_1$ and $\bar{C}_2$ respectively. The two functions

$$\phi_n'(z) - f_n''(z), \qquad \phi_n''(z) - f_n'(z)$$

are analytic in $\bar{C}_1$ and $\bar{C}_2$ respectively, and their difference approaches the value zero at each of the points $z = \alpha_i$. Let ϵ_n be the largest

$$\left| [\phi_n'(\alpha_i) - f_n''(\alpha_i)] - [\phi_n''(\alpha_i) - f_n'(\alpha_i)] \right|, \qquad \text{for } i = 1, 2, \cdots, k,$$

so that ϵ_n approaches zero as n becomes infinite. Then there exists* a rational function $\Phi_n(z)$, $n = 1, 2, \cdots$, analytic in $\bar{C}_2$ such that

$$\Phi_n(\alpha_i) = \phi_n'(\alpha_i) - f_n''(\alpha_i),$$
$$\left| \Phi_n(z) - [\phi_n''(z) - f_n'(z)] \right| < N\epsilon_n,$$

uniformly for z in $\bar{C}_2$, where N is a suitable number independent of n.

The function

$$\begin{cases} \phi_n'(z) - f_n''(z), & z \text{ in } \bar{C}_1, \\ \Phi_n(z), & z \text{ in } \bar{C}_2, \end{cases}$$

is analytic interior to C_1 and interior to C_2 and continuous on the entire closed point set $\bar{C}_1 + \bar{C}_2$, so by Theorem XIV there exists a rational function $r_n(z)$ such that we have

$$\left| r_n(z) - [\phi_n'(z) - f_n''(z)] \right| \leqq \epsilon_n, \qquad z \text{ in } \bar{C}_1,$$
$$\left| r_n(z) - \Phi_n(z) \right| \leqq \epsilon_n, \qquad z \text{ in } \bar{C}_2.$$

Then the rational function

$$R_n(z) = r_n(z) + f_n'(z) + f_n''(z)$$

satisfies the equation

---

* See the proof of Theorem X, loc. cit.

$$R_n(z) - [f_n'(z) + \phi_n'(z)] = r_n(z) - [\phi_n'(z) - f_n''(z)];$$

the right-hand member is in absolute value not greater than ϵ_n on $\bar{C}_1$, so that $R_n(z)$ approaches $f_1(z)$ as n becomes infinite, uniformly on $\bar{C}_1$ or on the sub-set of $\bar{C}_1$ on which $f_1(z)$ is defined. We have similarly

$$\begin{aligned} R_n(z) - [f_n''(z) + \phi_n''(z)] &= r_n(z) - [\phi_n''(z) - f_n'(z)] \\ &= [r_n(z) - \Phi_n(z)] + \{\Phi_n(z) - [\phi_n''(z) - f_n'(z)]\}, \end{aligned}$$

which is in absolute value not greater than $(N+1)\epsilon_n$ on $\bar{C}_2$, so that $R_n(z)$ approaches $f_2(z)$ as n becomes infinite, uniformly on $\bar{C}_2$ or on the sub-set on which $f_2(z)$ is defined. This completes the proof.

It will be noticed that we have actually proved more than is stated formally in the theorem, for we have shown

$$R_n(z) - [f_n'(z) + \phi_n'(z)] \to 0 \quad \text{uniformly for } z \text{ in } \bar{C}_1,$$

$$R_n(z) - [f_n''(z) + \phi_n''(z)] \to 0 \quad \text{uniformly for } z \text{ in } \bar{C}_2.$$

If the two Jordan curves C_1 and C_2 lie exterior to each other, having no point in common, and if $f_1(z)$ and $f_2(z)$ defined on $\bar{C}_1$ and $\bar{C}_2$ or on sub-sets thereof are uniformly developable on those point sets in series of rational functions, then there exists a sequence of rational functions which approaches $f_1(z)$ uniformly on $\bar{C}_1$ and $f_2(z)$ uniformly on $\bar{C}_2$ or on the point sets thereof. This can be proved easily by the methods of Runge, or indeed by means of Theorem XV.

We remark incidentally that Theorem XV enables us to prove Theorem VI by means of Theorem V; this remark is of some interest in virtue of the fact that the original proof of Theorem V is by means of Theorem VI.

7. **Proof of Theorem VIII.** As a matter of convenience we state

THEOREM XVI. *Let the regions $R_1, R_2, \cdots, R_n$, each bounded by a finite number of Jordan curves, be mutually exclusive, and let no two of the corresponding closed regions have more than a finite number of points in common. Then if the function $f_i(z)$ defined in $\overline{R}_i$ or more generally on a sub-set of $\overline{R}_i$ which contains the points common to $\overline{R}_i$ and the other regions $\overline{R}_k$, is in $\overline{R}_i$ (or on the sub-set) uniformly developable in a series of rational functions of z, for $i=1, 2, \cdots, n$, and if the function*

$$f(z) = f_i(z), \; z \text{ in } \overline{R}_i \text{ or on the prescribed sub-set of } \overline{R}_i,$$

is uniquely defined and continuous on $\overline{R}_1+\overline{R}_2+\cdots+\overline{R}_n$ or on the corresponding sub-set, then the function $f(z)$ is developable in a uniformly convergent series of rational functions of z on $\overline{R}_1+\overline{R}_2+\cdots+\overline{R}_n$ or on the sub-set of this latter set on which $f(z)$ is defined.

The proof is by induction; the case $n=2$ follows immediately from Theorem XV if $\overline{R}_1$ and $\overline{R}_2$ have common points, and from the method of Runge in the contrary case. Assume the theorem true for $n-1$ regions R_i; we shall prove it true for n such regions.

Let C be any Jordan curve which bounds completely or partially a region R_i and which separates two of the regions $R_1, R_2, \cdots, R_n$. Then there exists another Jordan curve C', which has no points in common with (the interior points of) $R_1+R_2+\cdots+R_n$, which has not more than a finite number of points in common with C, each of these a point of definition of $f(z)$, and such that no region R_i lies between C and C'. To the given regions interior to C (or exterior to C, if the interior of C contains all but a finite number of points of C') we apply the present theorem, for the number of regions concerned is less than n, and we proceed similarly for the given regions interior to C' (or exterior to C', if the interior of C' contains all but a finite number of points of C). Then we apply Theorem XV to the interior of C and the interior of C' (or to the exteriors in the respective cases described), the sub-sets here being the given regions or their sub-sets, which completes the proof in case C and C' actually possess common points. If C and C' do not possess common points, the method of Runge instead of Theorem XV yields the conclusion desired.

Theorem VIII now follows from Theorem XVI. The point set M consists of a finite number of closed regions, each bounded by a finite number of Jordan curves, and of a finite number of Jordan arcs A exterior to those regions; if M is composed partly of Jordan curves not bounding regions which belong to M, we consider each of these curves broken up into a finite number of Jordan arcs. The entire set of Jordan arcs may be considered, by a suitable change of notation if necessary, to have no points other than end points common to any two of them, or common to a Jordan arc and any one of the Jordan curves which bound regions belonging to M. Then each Jordan arc A can be enclosed in a new region bounded by a new Jordan curve whose interior (on the sphere) contains the interior points of A and no other points of M, and so that the new curve has no points in common with M or with other such new curves except such end points of A as belong likewise to other Jordan arcs or curves belonging to M.

These new Jordan regions, together with the regions originally belonging to M, form a set of regions to which Theorem XVI can be applied immediately; the given function $f(z)$ is of course not defined on the entire closed point set consisting of these regions, only on the sub-set M, but the possibility of approximating $f(z)$ on M follows at once.

The second part of Theorem VIII can be obtained by the following

reasoning. Any rational function continuous on M has its poles in the k regions into which M divides the plane, and can be separated into partial fractions each of which has merely a single pole. If R_1 denotes one of those k regions and z_1 the chosen point of R_1, any rational function with no poles outside of R_1 may, by Runge's theorem (i.e., Theorem I, after transformation $z'=1/(z-z_1)$ if necessary), be approximated on M as closely as desired by a rational function with but a single pole and that in z_1. Thus all component parts of the rational functions approximating to $f(z)$ can be approximated on M by rational functions with poles in the preassigned points, so that fact is true of $f(z)$ itself. In particular if $k=1$ and M does not contain the point at infinity, $f(z)$ can be approximated on M as closely as desired by a polynomial in z. This remark (second part of Theorem VIII) is indeed entirely general in its application, and refers to practically all the theorems of the present paper.

If the function $f(z)$ to be approximated is even or odd, and if the point set M is symmetric in the origin, then the function $f(z)$ can be uniformly approximated as closely as desired by even or odd rational functions. This remark is entirely general; we prove a broader result, that if $f(z)$ satisfies the functional equation

$$f(\omega z) \equiv \lambda \cdot f(z), \quad \text{where} \quad \omega^n = \lambda^n = 1,$$

and if ωz belongs to M whenever z belongs to M, then $f(z)$ can be uniformly approximated on M by rational functions which satisfy this same functional equation. Let us assume

$$|r(z) - f(z)| < \epsilon, \quad z \text{ on } M,$$

where $r(z)$ is rational. Then we have also

$$|r(\omega z) - f(\omega z)| < \epsilon, \quad z \text{ on } M,$$

that is,

$$|r(\omega z) - \lambda \cdot f(z)| < \epsilon, \quad z \text{ on } M.$$

By continued replacing of z by ωz we obtain

$$|r(\omega^2 z) - \lambda^2 f(z)| < \epsilon, \quad z \text{ on } M,$$

$$\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot$$

$$|r(\omega^{n-1} z) - \lambda^{n-1} f(z)| < \epsilon, \quad z \text{ on } M.$$

We define $\rho(z)$ by means of the equation

$$n\lambda^{n-1}\rho(z) \equiv \lambda^{n-1} r(z) + \lambda^{n-2} r(\omega z) + \cdots + \lambda \cdot r(\omega^{n-2} z) + r(\omega^{n-1} z);$$

it is seen by direct substitution that $\rho(z)$ satisfies the functional equation

$$\rho(\omega z) \equiv \lambda \cdot \rho(z).$$

Moreover we find from the inequalities already considered, by addition after multiplication by suitable powers of λ,

$$\left| n \cdot \lambda^{n-1} \rho(z) - n \cdot \lambda^{n-1} f(z) \right| < n\epsilon, \quad z \text{ on } M,$$

that is,

$$\left| \rho(z) - f(z) \right| < \epsilon, \quad z \text{ on } M,$$

so that the rational function $\rho(z)$ has the properties required.

Likewise if the function $f(z)$ to be approximated satisfies the functional equation $f(\bar{z}) \equiv \bar{f}(z)$, and if the point set M is symmetric in the axis of reals, the approximating rational functions can be chosen to satisfy this same functional equation. If we have

$$\left| f(z) - r(z) \right| < \epsilon, \quad z \text{ on } M,$$

where $r(z)$ is rational, then we have also

$$\left| f(\bar{z}) - r(\bar{z}) \right| < \epsilon, \quad z \text{ on } M,$$

and, by virtue of the equation $\bar{f}(\bar{z}) \equiv f(z)$,

$$\left| f(z) - \bar{r}(\bar{z}) \right| < \epsilon, \quad z \text{ on } M.$$

From the first and last of these inequalities we obtain

$$\left| f(z) - \frac{r(z) + \bar{r}(\bar{z})}{2} \right| < \epsilon, \quad z \text{ on } M;$$

the rational function $R(z) \equiv \frac{1}{2}[r(z) + \bar{r}(\bar{z})]$ satisfies the functional equation $\overline{R}(z) \equiv R(\bar{z})$. This determination of the new function $R(z)$ is independent of the previous functional equation considered; if we have $r(\omega z) \equiv \lambda r(z)$, then we have also $R(\omega z) \equiv \lambda R(z)$.

Theorem VIII (and also Theorem XVI) can obviously be slightly extended, to include functions $f(z)$ not continuous on M but meromorphic in the interior points of M and even with properly restricted infinite values on the boundary of M.

Theorem VIII has applications to conformal mapping analogous to results considered in detail for the polynomial case.* We remark too that Theorem X in conjunction with Theorem VIII gives a more precise result than Theorem VIII itself.

---

* Walsh, these Transactions, vol. 30 (1928), pp. 472–482, §7.

These remarks all have application as well to the results of the two succeeding paragraphs.

8. **An infinity of Jordan arcs.** The methods already used apply, with some modification, to certain point sets composed of an infinity of Jordan arcs, curves, and regions. We shall use the term *Jordan element* E to denote a single point, a Jordan arc, or a region bounded by a finite number of Jordan curves no two of which have more than a finite number of points in common. A *Jordan configuration* K is a point set consisting of a denumerable or non-denumerable infinity of Jordan elements of which no two have more than a finite number of points in common. The *limit configuration* L of a Jordan configuration K is the point set, necessarily closed, such that in each neighborhood of an arbitrary point of L lie points of an infinity of different Jordan elements E belonging to K. It is of course true that the limit configuration is not always uniquely defined by the Jordan configuration itself, but to some extent by the notation. We now prove

THEOREM XVII. *Let M be an arbitrary closed point set and N a Jordan configuration which consists of a set of points and of Jordan arcs whose limit configuration is contained in M. Then an arbitrary function $f(z)$ continuous on $M+N$ and which on M is uniformly developable in a series of rational functions of z, is likewise uniformly developable on $M+N$ in such a series.*

In particular if N is an arbitrary Jordan configuration which is a closed point set consisting of a set of points and of Jordan arcs whose limit configuration (or whose nth limit configuration) consists of a finite number of points and Jordan arcs which do not divide the plane into more than a finite number of regions, then an arbitrary function $f(z)$ continuous on N can be developed uniformly on N in a series of rational functions of z.

We use here the idea of *nth limit configuration*. The second limit configuration of N is defined as the limit configuration of the (first) limit configuration, and the other limit configurations are likewise found by iteration. With this understanding, the second part of Theorem XVII is by virtue of Theorem IX contained in the first part. We proceed to the proof of the first part.

We find it convenient for purposes of exposition to consider N to consist, except for possible points belonging to M, of a set of Jordan arcs. This is always possible, for a point P of N not a point of M nor a point of a Jordan arc belonging to N is isolated, and there can be drawn through P a Jordan arc having no point of M nor of N other than P itself in common with M or N. This can be accomplished, moreover, for every requisite point P without altering the limit configuration of N. The definition of the function

$f(z)$ can be extended over these new Jordan arcs so that the new function is continuous over the entire new point set.*

Let an arbitrary positive ϵ be given; there exists a rational function $r(z)$ which on M differs from $f(z)$ by less than ϵ. The function $r(z)$ has no poles on M, hence is continuous and uniformly continuous on a closed point set M' which contains M in its interior and which is bounded by a finite number of Jordan curves. The function $f(z)$ is also uniformly continuous on $M+N$. Suppose for convenience in exposition that $M+N$ does not contain the point at infinity, and suppose $\delta>0$ so chosen that $|z'-z''|\leqq\delta$ implies

$$|r(z')-r(z'')|<\epsilon, \quad z',\ z'' \text{ in } M',$$

$$|f(z')-f(z'')|<\epsilon, \quad z',\ z'' \text{ on } M+N.$$

Choose a point set M'' consisting of a finite number of regions, each bounded by a finite number of Jordan curves of which no two have a common point, so that M'' contains M in its interior but is contained in M', and such that no point of M'' is at a distance of more than δ from some point of M. At most a finite number of the Jordan arcs A composing N have points exterior to M''; otherwise we should have infinitely many points belonging to different Jordan arcs of N with a limit point exterior to M'', hence not in M. It is conceivable that one or more Jordan arcs belonging to N should have infinitely many points belonging to the boundary of M'', but we can and do avoid this possibility by diminishing the size of M'' if necessary.

We now define a function $f_1(z)$ continuous on $M''+N$ which differs but little from $f(z)$ on $M+N$. The function $f_1(z)$ shall be equal to $r(z)$ on M''. Let A be an arbitrary Jordan arc of N which has one or more points in common with M'' but which does not lie wholly within M''; we change the notation if necessary so that A has precisely one point, an end point P, in common with M''. Let Q be a point of A not a point of M'' at a distance from P not greater than δ, and such that no point of A on the arc PQ of A is at a distance from P greater than δ. Define the function $f_1(z)$ equal to $r(z)$ at P, equal to $f(z)$ at Q, and varying monotonically† along A from P to Q. The function $f_1(z)$ differs from $f(z)$ at most by 3ϵ in P, for there exists a point R of M at a distance not greater than δ from P, and we have $f_1(P)=r(P)$,

---

* Lebesgue, Rendiconti del Circolo Matematico di Palermo, vol. 24 (1907), pp. 371–402; p. 379.

† That is, the function $w=f_1(z)$ maps the Jordan arc PQ one-to-one and continuously on a line segment in the w-plane.

$$|r(P) - r(R)| < \epsilon,$$
$$|r(R) - f(R)| < \epsilon,$$
$$|f(R) - f(P)| < \epsilon.$$

Hence the function $f_1(z)$ differs from $f(z)$ at most by 5ϵ along the arc PQ. The function $f_1(z)$ is defined equal to $f(z)$ on the portions of arcs A exterior to M'' but not arcs PQ. The function $f_1(z)$ is likewise defined equal to $f(z)$ on the Jordan arcs which belong to N but have no points in common with M''.

The function $f_1(z)$ is continuous on $M''+N$ and hence by Theorem VIII or Theorem XVI can be approximated on $M''+N$ by a rational function of z with an error not greater than ϵ, so $f(z)$ can be approximated on $M+N$ by a rational function of z with an error not greater than 6ϵ, and the theorem is established.

Many point sets more general than Jordan curves, or indeed more general than boundaries of regions, are included in the hypothesis of the theorem, so we have (compare §1) a new generalization of Weierstrass's theorem on approximation by trigonometric polynomials. For instance we may consider the region bounded by the point set

$$y = \sin\frac{1}{x}, \qquad -\frac{1}{\pi} \leqq x \leqq \frac{1}{\pi};$$
$$x = -\frac{1}{\pi}, \qquad -2 \leqq y \leqq 0;$$
$$x = \frac{1}{\pi}, \qquad -2 \leqq y \leqq 0;$$
$$x = 0, \qquad -1 \leqq y \leqq 1;$$
$$y = -2, \qquad -\frac{1}{\pi} \leqq x \leqq \frac{1}{\pi}.$$

Any function continuous on this closed point set (i.e., boundary of the region) can be uniformly approximated on this point set as closely as desired by polynomials in z and $1/(z+3i/2)$.

We may consider likewise a region consisting of a strip closed at one end which lies exterior to the unit circle, and which approaches this circle by winding about it infinitely many times. An arbitrary function continuous on the boundary of this region can be uniformly approximated on that boundary by rational functions of z, with poles only in the origin, at infinity, and at an arbitrarily chosen point interior to the region.

Similarly consider the point set $x=n$, $y=n$, where n takes all integral values, positive, negative, or zero. An arbitrary function continuous (in the extended plane) on this point set can be uniformly approximated on the point set by rational functions of z.

9. **An infinity of Jordan or more general regions.** The extension of Theorem VIII to include infinitely many *regions* is not so simple:

THEOREM XVIII. *Let infinitely many regions R form a Jordan configuration whose limit configuration L is contained in a closed point set M. Suppose moreover that the totality of regions R not belonging to M have at most a finite number of distinct points in common with M, and suppose that if an arbitrary region Q be considered which contains M in its interior, then each of those regions R not lying wholly in Q has at most a finite number of its points in common with other regions R, whether these other regions R lie wholly in Q or not.*

If $f(z)$ is a function which can be uniformly approximated on M by a rational function of z; if $f(z)$ is analytic interior to each region R or more generally if in each R or on a point set contained in each R but which includes all points common to this region R and either to M or to the other regions R—if on such a point set the function $f(z)$ can be uniformly approximated by a rational function of z; and if $f(z)$ is continuous on the entire closed point set consisting of M and the regions R (or the sub-sets), then on this closed point set the function $f(z)$ can be uniformly approximated as closely as desired by a rational function of z.

In particular if $M=L$ satisfies the condition given and consists of a finite number of points and of Jordan arcs which do not divide the plane into infinitely many regions, then an arbitrary function $f(z)$ analytic interior to each R and continuous on the entire point set $L+\{R\}$ can be uniformly approximated on that point set as closely as desired by a rational function of z.

The latter part of the theorem is an immediate consequence of the former part, in conjunction with Theorem IX. We proceed to consider the former part of the theorem.

The proof here is quite similar to the proof of Theorem XVII. Let an arbitrary positive ϵ be given, and let $r(z)$ be a rational function which differs from $f(z)$ on M by less than ϵ, and which coincides with $f(z)$ in the points common to M and the regions R (compare Theorem X). Assume neither M nor the regions R to contain the point at infinity, choose M' a closed point set containing M in its interior but containing on or within it no point of discontinuity of $r(z)$, and choose $\delta>0$ so that $|z'-z''|\leqq\delta$ implies

$$|r(z')-r(z'')|<\epsilon, \quad z', \quad z'' \text{ in } M',$$
$$|f(z')-f(z'')|<\epsilon, \quad z', \quad z'' \text{ on } L+\{R\}.$$

Choose a point set M'' containing M but contained in M', bounded by a finite number of Jordan curves of which no two have more than a finite number of common points, and such that no point of M'' is at a distance greater than δ from some point of M. It follows that for points of M'' where $f(z)$ is defined, the functions $r(z)$ and $f(z)$ differ by less than 3ϵ. We now modify M'', cutting away part of this point set and giving us a new point set M'', as follows. At most a finite number of regions R lie partly or wholly outside M''. A region R that lies partly outside M'' but has no points in common with M has Jordan arcs drawn, exterior to R, not meeting M nor other regions R, except that these Jordan arcs may have such points in common with R as R itself has with other regions R, modifying M'' so that the new M'' shall be bounded partly by these new Jordan arcs, and so that R and the new M'' shall have at most a finite number of common points, namely those points already explicitly mentioned. We have thus pared off a portion of M'' by means of Jordan arcs, so that except possibly for a finite number of points, R is isolated from M''. For the regions R which lie partly outside M'' but have points in common with M, we proceed similarly, cutting out parts of M'' adjacent to these regions by means of Jordan arcs, so that R and the new M'' have no common points other than those common to R and M and those common to R and other regions R.* The regions cut out of M'' are open regions, finite in number, each bounded by a finite number of Jordan arcs, containing in their interiors no point of M nor of a region R, and whose boundaries have at most a finite number of points in common with M, namely those points common to M and the regions R.

The point set consisting of the new M'' plus the regions R is then of precisely the kind considered in Theorem XVI; the function to be developed (or approximated) is $r(z)$ on M'' and $f(z)$ in the regions R or on the sub-sets, which is continuous on the entire point set considered. Theorem XVIII follows at once.

Iteration of this theorem gives (as in Theorem XVII) much more general results than the theorem itself.

We give merely a few of the very simplest examples to illustrate this theorem. In each case we give the open point set. A function analytic on the point set given and continuous on the corresponding closed point set can be

* One essential point is omitted here, namely the demonstration that *if a closed Jordan region R has but a finite number of points P in common with an arbitrary closed point set M, then there exists a second Jordan region R' whose boundary has only the points P in common with the boundary of R and precisely the same points in common with M, and whose interior (on the sphere) contains the interior of R but no points of M.* There is no difficulty about the proof if M is a Jordan region; in any case the proof can be given with comparative ease and is left to the reader.

uniformly approximated on the closed point set as closely as desired by a rational function of z.

(i) $$\left(x-\frac{1}{3^k}\right)^2+y^2<\frac{1}{4\cdot 3^{2k}} \qquad (k=0,1,2,\cdots);$$

(ii) $$x^2+2^{4k+2}\left(y-\frac{1}{2^{2k}}\right)^2<1 \qquad (k=0,1,2,\cdots);$$

(iii) $$3\pi/2^k<\arg z<5\pi/2^k \qquad (k=2,3,4,\cdots);$$

(iv) $$3/2^k<|z|<5/2^k \qquad (k=0,1,2,\cdots).$$

It is not to be supposed, however, that an arbitrary function analytic on an *arbitrary* point set and continuous on the corresponding closed point set can be uniformly approximated as closely as desired on that closed point set by means of rational functions of z. In fact there exists a function $f(z)$ not a constant but continuous over the entire plane, and analytic everywhere (even at infinity) except on the points of a Jordan arc C.* Let K be a circle which contains C in its interior. Then $f(z)$ is analytic in the interior points of K not points of C and is continuous on the corresponding closed point set, but on this closed point set cannot be uniformly approximated as closely as desired by a rational function of z. For an approximating rational function must be continuous in $\overline{K}$, must have its poles exterior to $\overline{K}$, hence leads directly to uniform approximation by a polynomial in z. Such approximation leads in turn to a series of polynomials convergent to the sum $f(z)$ uniformly on and within K, which means that $f(z)$ is analytic throughout the interior of K, so that $f(z)$ is a constant, which is a contradiction.

10. **Non-uniform expansion.** We devote a few lines to the very simplest applications of our results to the expansion, not necessarily uniform, of arbitrary functions in terms of rational functions. The following theorem is essentially contained in Runge's classical results:

Let M be an arbitrary open point set. Then an arbitrary function $f(z)$ meromorphic in each point of M can be expanded on M in a series of rational functions of z. This series converges uniformly on any closed point set belonging to M, and all poles of the functions of the series, except the poles of $f(z)$ itself, can be chosen to lie on the boundary of M, or indeed in arbitrarily chosen points, at least one in each of the regions or other point sets into which M divides the plane.

---

* This arc C is to be chosen so as to have positive area but is otherwise arbitrary. See Pompeiu, American Journal of Mathematics, vol. 32 (1910), pp. 327–332.

Consider, in fact, a sequence of closed point sets $M_1, M_2, \cdots$, constructed perhaps by the method of continued subdivision of the plane by a network of squares, such that each point set M_i is composed of a finite number of non-intersecting closed regions each bounded by a finite number of non-intersecting Jordan curves, such that each point set M_i lies interior to its successors and to M, and such that every point of M lies in some M_i. There exists a rational function $r_k(z)$ such that we have

$$| f(z) - r_k(z) | < \frac{1}{k}, \quad z \text{ on } M_k ;$$

all the poles of $r_k(z)$ except the poles of $f(z)$ may be chosen on the boundary of M. For each of the finite number of regions into which M_k separates the plane either contains a boundary point of M (compare Theorem VIII) or can itself be adjoined to M_k. The sequence $\{r_k(z)\}$ leads directly to the required series $r_1(z) + [r_2(z) - r_1(z)] + \cdots$. It is an essential part of the proof to notice that if M is the entire plane, the function $f(z)$ itself must be rational, and all the functions $r_k(z)$ can be chosen equal to $f(z)$.

The method of proof just used yields also the following:

Let the point set M be the sum of the open sets $M_1, M_2, \cdots$, no two of which have a common point, and suppose the function $f_i(z)$ can be expanded on the point set M_i or on a subset thereof in a sequence of rational functions $\{r_{ij}(z)\}$. Then these functions $f_i(z)$ can be expanded on M or on the corresponding subset in a single sequence of rational functions $\{r_j(z)\}$; if M' is a closed set belonging to M, and if on the subset of M' in each M_i the sequence $\{r_{ij}(z)\}$ converges uniformly, then the sequence $\{r_j(z)\}$ converges uniformly on M'.

In particular some or all of these point sets M_i may be chosen as regions of arbitrary connectivity, and the functions $f_i(z)$ corresponding as arbitrary functions meromorphic in those regions; the sequence $\{r_j(z)\}$ converges uniformly on any closed point set interior to the totality of such M_i.

Let the point set M_i be expressed as the sum of the point sets M_{ij}, so that M_{ij} consists of a finite number of non-intersecting closed regions each bounded by a finite number of non-intersecting Jordan curves, so that M_{ij} lies interior to M_i and to $M_{i,j+1}$, and so that each point of M_i lies in some M_{ij}. The sequence $\{r_j(z)\}$ can be defined so that we have

$$| r_{ij}(z) - r_j(z) | < \frac{1}{j}, \quad z \text{ on } M_{1j} + M_{2,j-1} + \cdots + M_{j1}.$$

The uniformity of the convergence as stated follows from the fact that

no closed set belonging to M can contain points of an infinity of the point sets M_i.

We establish next a theorem concerning Jordan arcs, more specific than a similar theorem on expansion in series of polynomials recently proved by Hartogs and Rosenthal.*

Let the point set C consist of an infinity of Jordan arcs $C_1, C_2, \cdots$, of which no two have more than a finite number of points in common. Then an arbitrary function $f(z)$ continuous on C can be expanded on C in a series of rational functions of z, and the series converges uniformly on each arc C_k.

The proof is entirely obvious; the series is formed from the sequence $\{r_k(z)\}$, where $r_k(z)$ is so determined that we have

$$|f(z) - r_k(z)| < \frac{1}{k}, \quad z \text{ on } C_1 + C_2 + \cdots + C_k;$$

this choice of $r_k(z)$ is possible by Theorem IX. In particular if no finite set of arcs $C_1+C_2+\cdots+C_k$ divides the plane, and if C does not contain the point at infinity, it is clear by the second part of Theorem VIII that the rational function $r_k(z)$ can be chosen as a polynomial in z.

We establish now a result more general than those just considered; a somewhat less general result has recently been given by Carathéodory.†

Let the point set M be composed of the mutually exclusive sets $D_0, D_1, D_2, \cdots$, of which the first is open (and may be empty) and the others closed, and suppose the function $f_i(z)$ can be expanded on D_i or on a subset thereof in a sequence of rational functions $\{r_{ij}(z)\}$. Then those functions $f_i(z)$ can be expanded on M or on the corresponding subset in a single sequence of rational functions $\{r_k(z)\}$.

The sequence $\{r_k(z)\}$ converges uniformly on any closed set belonging to D_0 on which the sequence $\{r_{0j}(z)\}$ converges uniformly, and on any subset of D_i, $i>0$, on which $\{r_{ij}(z)\}$ converges uniformly.

In particular some or all of these point sets D_i, $i>0$, may be chosen as closed regions of arbitrary connectivity, and the functions $f_i(z)$ corresponding as arbitrary functions meromorphic in the closed regions. Some or all of the D_i, $i>0$, may be chosen as boundaries of regions composing D_0, or as boundaries of other regions, and the corresponding functions $f_i(z)$ as arbitrary functions meromorphic on those point sets; the sequence $\{r_k(z)\}$ can be chosen uniformly convergent on each such D_i. If some of the sets D_j are

* Mathematische Annalen, vol. 100 (1928), pp. 212–263.

† Bulletin of the American Mathematical Society, vol. 34 (1928), pp. 721–725.

composed of a finite number of Jordan arcs, no two having more than a finite number of common points, the corresponding functions $f_j(z)$ can be chosen as arbitrary continuous functions and the sequence $\{r_k(z)\}$ can be made to converge uniformly on each such D_j. If a point set D_j is composed of Jordan arcs, the function $f_j(z)$ may be chosen as an arbitrary function of Baire's class 0 or 1; in this case the convergence on D_j of the sequence $\{r_k(z)\}$ is not necessarily uniform.

We proceed to the proof of the theorem. We construct, perhaps by the same method previously employed of continued subdivision of the plane into squares, a double sequence of closed point sets D_{ij}. The point sets D_{ij} shall each consist of a finite number of mutually exclusive closed regions each bounded by a finite number of non-intersecting Jordan curves. The set D_{0k} shall lie completely interior to $D_{0,k+1}$ and to D_0, and each point of D_0 shall lie interior to some set D_{0k}. The point sets D_1 and $D_{1,k+1}$ shall lie interior to D_{1k}, and no point not in D_1 shall lie in all the D_{1k}. Moreover, D_{1k} is to be chosen so as to have no point in common with $D_{0,k+1}$. The remaining point sets are to be similarly constructed; the sets D_i and $D_{i,k+1}$ shall lie interior to D_{ik}, and no point not in D_i shall lie in all the D_{ik}. Moreover, D_{ik} is to be chosen so as to have no point in common with $D_{0,i+k}$, $D_{1,i+k-1}$, $\cdots$, $D_{i-1,k+1}$. The sequence of rational functions $\{r_n(z)\}$ required may now be determined by the conditions

$$
\begin{aligned}
&|r_n(z) - r_{0n}(z)| < 1/n, \quad z \text{ on } D_{0n},\\
&|r_n(z) - r_{1,n-1}(z)| < 1/n, \quad z \text{ on } D_{1,n-1},\\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\
&|r_n(z) - r_{n-1,1}(z)| < 1/n, \quad z \text{ on } D_{n-1,1}.
\end{aligned}
$$

The paper by Hartogs and Rosenthal to which reference has already been made considers the general problem of *uniform or non-uniform* expansion of arbitrary functions, and by means of polynomials instead of by more general rational functions. There is thus comparatively little overlapping between that paper and the present one. We mention, however, the following theorem, proved by Hartogs and Rosenthal for the more restricted case of polynomials, which can be proved in its present form by a combination of the methods given by them and those of the present paper:

If M is an arbitrary closed point set and if $f(z)$ is an arbitrary function continuous on M such that on each component of M the function $f(z)$ can be approximated as closely as desired by a rational function of z, then $f(z)$ can be approximated as closely as desired on the entire point set M, by a rational function of z.

The word *component* (i.e. *Stück*) here means a continuum (or single point) belonging to M but not contained in any other continuum belonging to M.

11. New proof of Theorem XI. It is perhaps of interest to remark that Theorem XI can be proved by the methods of the present paper. Inspection of the proof about to be given and of the proof of Theorem XII shows that it is sufficient in our present proof if we take only the special case that two circles exist externally tangent to each other at $z=\alpha$, neither of which contains a point of $\bar{C}_1$ or $\bar{C}_2$ in its interior. We suppose too that in the neighborhood of α the curves C_1 and C_2 lie on opposite sides of the line of centers of these two circles. Then there exists a Jordan curve C_1', with continuously turning tangent and continuous curvature except at $z=\alpha$ and at a second point $z=\beta_1$, which contains all points interior to C_1 but no point interior to C_2 in its interior, which has no point other than α in common with $\bar{C}_2$, which in the neighborhood of α consists of an arc of each of the two circles mentioned, and which is of such a nature that at β_1 a Jordan arc A_1 can be joined on exterior to C_1' so that each of the two arcs of C_1' from α to β_1 form with A_1 an arc with continuously turning tangent and continuous curvature. We find too a second Jordan curve C_2' which has analogous properties with respect to C_2, and we join the two arcs A_1 and A_2 so that they together form a Jordan arc A which has a continuously turning tangent and continuous curvature, and which has no point other than β_1 and β_2 in common with C_1' and C_2'.

By using such a transformation as $w^2=z-\beta$, we find as in the proof of Theorem XII a function $f_1(z)$ analytic in $\bar{C}_1'$ except at $z=\alpha$, and differing from $f(z)$ in $\bar{C}_1$ by less than an arbitrary preassigned positive ϵ. We find likewise a function $f_2(z)$ analytic in $\bar{C}_2'$ except at $z=\alpha$ and differing from $f(z)$ in $\bar{C}_2$ by less than ϵ, and can assume $f_1(\alpha)=f_2(\alpha)=f(\alpha)$. Denote by Γ_1 and Γ_2 two Jordan curves which have in common the point $z=\alpha$ and the arc A, and which are each composed (in addition to A) of an arc of C_1' and an arc of C_2'. We assume $\bar{C}_1$ and $\bar{C}_2$ to lie interior to Γ_1 and exterior to Γ_2. Extend the definition of the function*

$$F(z) = \begin{cases} f_1(z), & z \text{ on } \bar{C}_1', \\ f_2(z), & z \text{ on } \bar{C}_2', \end{cases}$$

* The function $F(z)$ is easily found in the w-plane, $w^2=z-\beta$, where β is interior to Γ_1 and Γ_2. Approximate by a polynomial in w the function $f_1(z)$ in the transform of the region $\bar{C}_1$, and the function $f_2(z)$ in the transform of the region $\bar{C}_2$, so that this polynomial takes on the same value in the two points $w=\pm(\alpha-\beta)^{1/2}$. The transform in the z-plane of this polynomial in w may be used as the function $F(z)$.

so that $F(z)$ is defined over Γ_1 and Γ_2 and satisfies a Lipschitz condition on both of these curves.

Cauchy's Integral for the function $F(z)$ (and for z in C_1 or C_2 this integral formula is valid) may be written

$$F(z) = \frac{1}{2\pi i}\int_{C_1'} \frac{F(t)dt}{t-z} + \frac{1}{2\pi i}\int_{C_2'} \frac{F(t)dt}{t-z}$$
$$= \frac{1}{2\pi i}\int_{\Gamma_1} \frac{F(t)dt}{t-z} - \frac{1}{2\pi i}\int_{\Gamma_2} \frac{F(t)dt}{t-z},$$

where the integrals are taken in the positive (i.e., counterclockwise) sense over the curves indicated. By Plemelj's theorem, the first integral in the right-hand member represents a function $\phi_1(z)$ analytic in Γ_1 and continuous in the corresponding closed region, and the second integral represents a function $\phi_2(z)$ analytic exterior to Γ_2 and continuous in the corresponding closed region.

There exists a Jordan region J_1 within which $\phi_1(z)$ is analytic and which contains every interior point of C_1 and of C_2 in its interior and such that $\phi_1(z)$ is continuous in the closed region $\bar{J}_1$. In $\bar{J}_1$ the function $\phi_1(z)$ can be approximated uniformly by a polynomial in z. There exists likewise a Jordan region J_2 within which $\phi_2(z)$ is analytic and which contains every interior point of C_1 and of C_2 in its interior and such that $\phi_2(z)$ is continuous in the closed region $\bar{J}_2$. In $\bar{J}_2$ the function $\phi_2(z)$ can be uniformly approximated by a polynomial in z. Hence in $\bar{C}_1+\bar{C}_2$ the function $F(z)$ can be uniformly approximated as closely as desired by a polynomial in z, so the same is true of the function $f(z)$, and Theorem XI is established.

HARVARD UNIVERSITY,
CAMBRIDGE, MASS.

SEPARATION THEOREMS WITH APPLICATIONS TO QUESTIONS CONCERNING ACCESSIBILITY AND PLANE CONTINUA*

BY

R. G. LUBBEN

In this paper we shall confine ourselves to a two-dimensional euclidean space which we shall denote by the symbol S.

R. L. Moore† has given the following extension of a theorem by Zoretti:‡ *If K is a bounded maximal connected subset of a closed point set M and does not separate space, and e is a positive number, then there exists a simple closed curve containing no point of M, such that the interior of this curve contains K but contains no point whose distance from K is greater than e.* Zoretti's conclusion is weaker than Moore's in that it says nothing about the points on the *interior* of the curve surrounding K. On the other hand, Moore finds it necessary to use a stronger hypothesis than Zoretti's; namely that K does not separate the plane. In this paper we consider among other things the question of the conditions involved in a combination of Zoretti's hypothesis and Moore's conclusion, and show in Theorem 1 that a result analogous to

---

* Various parts of this paper were presented to the American Mathematical Society on September 10, 1925, December 30, 1925, and February 27, 1926; received by the editors August 31, 1928.

† Cf. R. L. Moore, *Concerning the separation of point sets by curves*, Proceedings of the National Academy of Sciences, vol. 11 (1925), pp. 469–476.

‡ Cf. L. Zoretti, *Sur les fonctions analytiques uniformes*, Journal de Mathématiques Pures et Appliquées, (6), vol. 1 (1905), pp. 9–11. A pair of point sets M_1 and M_2 are said to be *mutually separated* provided that neither contains a point or a limit point of the other. A point set is said to be *connected* if it is not the sum of two non-vacuous mutually separated point sets. A point set M is said to be *strongly connected* provided that for every pair of points belonging to M there exists a closed, connected subset of M containing these points; cf. R. L. Moore, *Concerning continuous curves in the plane*, Mathematische Zeitschrift, vol. 15 (1922), p. 254. If H is a point set and M is a connected subset of H such that every connected subset of H which has a point in common with M is a subset of M, then M is said to be *a maximal connected subset of H*. A point set M is said to *separate space* if $S-M$ is not connected. A *simple continuous arc AXB*, whose *end points* are the distinct points A and B, is a closed, bounded, connected set of points, M, containing A and B such that if P is a point of $M-A-B$, $M-P$ is not connected; cf. R. L. Moore, *Concerning simple continuous curves*, these Transactions, vol. 21 (1920), p. 340. The point set $M-A-B$ mentioned above is the *segment AXB* of the simple continuous arc AXB. For definitions of a simple closed curve, of an open curve, and of other of the simpler point sets see the paper mentioned above, the paper F. A., and articles to which references are made in these two papers. We shall use the notation F. A. to denote R. L. Moore's paper, *On the foundations of plane analysis situs*, these Transactions, vol. 17 (1916), pp. 131–164.

that of Moore holds, if we allow the boundary of the domain covering K to consist of a finite number of simple closed curves. In Theorems 2 and 3 we give conditions under which it is possible to cover a bounded point set by a finite number of such domains.

It follows from these theorems that a given point in one of two closed, bounded, mutually exclusive point sets can be separated from a given point in the other by a simple closed curve containing no point of the sum of the two point sets. In problems concerning the boundaries of domains, accessibility, and the separation of unbounded point sets by curves, it is of interest to consider the question of the separation of sets, having points in common, by simple closed curves containing no points of the given sets, except those that are common to them. R. L. Moore* has given sufficient conditions for the existence of such a curve, for the case where both point sets are bounded continua. We generalize his results by giving conditions which are both necessary and sufficient, and by removing the condition that both point sets be continua; see Theorems 16 and 17. Theorem 18 is concerned with the separation of a disconnected subset of a continuum by a simple closed curve. The case where the point set T mentioned in this theorem consists of one point is of particular interest, since it implies the existence of separation curves, for the case of a continuum which is disconnected by the omission of one of its points; see Theorems 21 and 23. Theorems such as 21 and 23 are often useful in proving the connectivity of point sets. The case where T consists of one or two points is of particular interest in questions concerning the separation of unbounded point sets; see the statements and proofs of Theorems 20, 21, and 23.

We find useful the notion of one point set's being connected near another, and the notion of one point set's not being separated by another near a third point set. These concepts play a fundamental role in our treatment of separation theorems: in Theorems 6, 7, and 9 concerning conditions sufficient to make a point set strongly connected, and Theorems 12, 14, and 15 concerning the relation of a domain to its boundary. Theorems 4, 11, 13, 22, and 24 are concerned with conditions for accessibility.

THEOREM 1. *If M is a bounded maximal connected subset of a closed point set K and e is a positive number, then there exists a multiply connected Jordan domain H containing M such that* (1) *the boundary of H contains no point of K, and* (2) *if P is a point of H', then $d(P, M) < e$.*†

* See *Concerning the separation of point sets by curves*, loc. cit., Theorem 2.

† If M is a point set, $S-M$ is the *complement* of M. A *domain* is a connected point set whose complement is closed. A *Jordan domain* or *Jordan region* is the interior of a simple closed curve. A

Since M is bounded, there exist at most a finite number, n, of bounded, mutually exclusive complementary domains* $D_1, D_2, D_3, \cdots, D_n$ of M which contain points whose distance from M is greater than $\frac{1}{2}e$. Let D_0 be the unbounded complementary domain of M. For each integer i ($i=0, 1, 2, 3, \cdots, n$), let C_i be a circle, which, with its interior, is a subset of D_i; let P_i be the center of C_i and L_i be a circle having the same center as C_i and having an interior containing M. Let N_i be the annular domain bounded by C_i+L_i. Then N_i contains M. Let the equation $w_0=f_0(z)$ define the identity transformation of the plane into itself. For $i=1, 2, 3, \cdots, n$, let the equation $w_i=f_i(z)$ define an inversion of the plane about the point P_i with reference to the circle C_i. If Q is a point set in the z plane, let $(Q)_i=f_i(Q)$.

For $i=0, 1, 2, 3, \cdots, n$, $z=f_i^{-1}(w_i)$ is a uniformly continuous function of w_i over the closed and bounded point set $(N_i)_i'$. Hence, there exists a positive number, d_i, such that when $|w_{1_i}-w_{2_i}|<d_i$, $|z_1-z_2|<\frac{1}{2}e$, where w_{1_i} and w_{2_i} are points of $(N_i)_i'$. Let m_i be the point set consisting of $(M)_i$ plus all bounded complementary domains of $(M)_i$. Let $k_i=m_i+(K)_i$.

If *m_i is not a maximal connected subset of k_i*, there must exist a connected subset of k_i which is the sum of m_i and t_i, where t_i is a point set containing no point of m_i. It is easily seen that both k_i and m_i are closed. Hence, if T_i is a maximal connected subset of t_i', it must have a limit point in m_i. Since T_i is a subset of the unbounded complementary domain of the closed point set $(M)_i$, T_i' must be a subset of $(K)_i$. Hence, $(M)_i+T_i$ is a connected subset of $(K)_i$. If T_i is non-vacuous, then M is not a maximal connected subset of K, contrary to the hypothesis of the theorem. Thus, k_i is a closed point set, and m_i is a bounded maximal connected subset of k_i, which does not separate space. Hence, there exists† a simple closed curve j_i, which encloses m_i, contains no point of k_i, and whose interior contains no point whose distance from m_i is greater than d_i. From the definition of j_i it follows that j_i is a subset of $(D_i)_i$, and that no point of $(D_i)_i$ within j_i is at a distance greater than d_i from $(M)_i$. Let $J_i=f_i^{-1}(j_i)$. Then the domain H bounded by the finite collection $\sum_{i=0}^{n} J_i$ satisfies the conclusion of the theorem.

---

multiply connected Jordan domain, of order n, is a domain whose boundary consists of n mutually exclusive simple closed curves. In this paper we shall consider only domains of a finite order. Thus,a Jordan region is of order unity. If M is a point set, by M' is meant the set of all limit points of M and $M+M'$ is denoted by $\overline{M}$. If M and N are point sets, by $d(M, N)$, *the distance from M to N*, we mean the lower bound of the distances between pairs of points, x and y, where x is a point of M and y of N.

* If M is a point set, by a *complementary* domain of M is meant a maximal connected subset of $S-\overline{M}$.

† Cf. Theorem 1 of Moore's paper *Concerning the separation of point sets by curves.*

THEOREM 2. *If K is a point set, e is a positive number, and L is a bounded point set which is either K or a closed point set which is the sum of a collection of maximal connected subsets of $\overline{K}$, then there exists a finite collection of multiply connected Jordan domains covering L and such that* (1) *the upper distance* of any of these domains from the product of L and that domain is less than e,* (2) *the distance between any pair whatever of domains of this collection is a positive number,* (3) *the boundary of this collection of domains contains no point of $\overline{K}$, and* (4) *each domain contains a point of L.*†

For each maximal connected subset A of $\overline{L}$ there exists, by Theorem 1, a multiply connected Jordan domain H_A such that no point of the boundary of this domain is a point of $\overline{K}$, and the upper distance of H_A' from its product with $\overline{L}$ is less than $e/6$. By the Heine-Borel-Lebesgue Theorem there exists a finite sub-collection H_1, which covers $\overline{L}$, of the collection $[H_A]$. Let T be the set of all points which belong to elements of H_1. There exist then a finite number, k, of maximal connected subsets, $T_1, T_2, T_3, \cdots, T_k$, of T'. Let d be a positive number which is smaller than $e/6$, and is also smaller than the smallest of the positive numbers $d(T_i, T_j)/3$, where $i, j = 1, 2, 3, \cdots, k$, and $i \neq j$. By an argument analogous to that used in a similar connection in the proof of Theorem 1 it can be shown that T_i $(i = 1, 2, 3, \cdots, k)$ is a maximal connected subset of $\overline{T}+\overline{K}$; hence, by Theorem 1, there exists a multiply connected Jordan domain N_i covering T_i such that no point of N_i is at a distance from T_i greater than d, and that the boundary of N_i contains no point of $\overline{T}+\overline{K}$. Then $\sum_{i=1}^{k} N_i$ is a collection of domains satisfying the conclusion of the theorem.

THEOREM 3. *If K is a closed point set, L is a closed and bounded point set which is the sum of a collection of maximal connected subsets of K, d is a positive number such that no maximal connected subset of L has a diameter greater than d, and e is any positive number whatever, then there exists a finite collection of Jordan domains covering L such that* (1) *no domain in this collection has a diameter greater than $e+d$;* (2) *if D and $\overline{D}$ are any two domains whatever in this collection, then $d(D, \overline{D}) > 0$;* (3) *the boundaries of these domains have no*

* By *the upper distance from M to N*, written $u(M, N)$, we mean the *upper bound* of the set of values $[d(A, N)]$, where A is a variable point of M. Cf. R. L. Moore, *Concerning upper semi-continuous collections of continua*, these Transactions, vol. 27 (1925), p. 416, second footnote. A collection M of domains is said to *cover* a point set K, provided every point of K belongs to some element of the collection M. The *diameter* of the point set M is the upper bound of the set of values $[d(A, B)]$ where A and B are points of M.

† Theorems 2 and 3 are generalizations of results stated by Schoenflies and Kerékjárto. Cf. A. Schoenflies, *Die Entwickelung der Lehre von den Punktmannigfaltigkeiten*, Berlin, Teubner, 1908, pp. 104–106; and B. von Kerékjárto, *Vorlesungen ueber Topologie*, Berlin, Springer, 1923, pp. 49–52

points in common with K; (4) each domain contains a point of L. In particular, if the point set L is totally disconnected,† the conclusion holds, if the number d mentioned in the hypothesis is zero.

The interior of a simple closed curve has the same diameter as the curve itself. It follows by Theorem 2 that if the theorem is not true, there must exist a positive number f and a sequence of positive numbers, $e_1, e_2, e_3, \cdots, e_n, \cdots$, such that (1) e_n approaches zero as n approaches infinity; (2) for each positive integer n there exists a finite collection H_n of multiply connected Jordan domains which have, with respect to the point sets L and K and the positive number e_n, the properties (1), (2), (3), and (4) mentioned in the conclusion of Theorem 2; (3) for each n the collection H_n contains a domain h_n of diameter greater than $d+f$. It can readily be shown that the collection $[h_n]$ has an infinite sub-collection which has a closed and connected limiting set,‡ and that this limiting set has a diameter not less than $d+f$. Since L is closed, this limiting set is a subset of L. Hence the supposition that the theorem is not true leads to a contradiction of the hypothesis that L has no connected subset of diameter greater than d.

THEOREM 4. *Given that A and B are distinct points and G is a bounded collection of point sets such that (1) if e is a positive number, there exist at most a finite number of elements of G having a diameter greater than e; (2) if G^* is the sum of the elements of G, and g is an element of G, then $d(g, G^*-g)>0$; and (3) for each element g of G there exists a simple continuous arc whose end points are A and B, and which contains no point in common with the point set g. Then there exists a simple continuous arc AWB which contains no point in common with G.**

It follows from condition (2) of the hypothesis that the elements of G are countable and hence can be put in a one-to-one correspondence with the set of positive integers. Let g_1 be the first element of G in this ordering, and let h_1 be a simple continuous arc which contains no point of g_1 and whose end points are A and B. For i an integer greater than unity, let g_i be the first

† A point set is said to be *totally disconnected* if it has no connected subset containing more than one point.

‡ Cf. S. Janiszewski, *Sur les continus irréductibles entre deux points*, Journal de l'École Polytechnique, (2), vol. 16 (1912), p. 97. By the *limiting set* of a collection G of point sets we mean the set of all points P such that every domain containing P contains points of infinitely many elements of the collection G. Cf. Janiszewski, loc. cit., p. 93, last four lines; A. Schoenflies, *Beiträge zur Theorie der Punktmengen*, Mathematische Annalen, vol. 59 (1904), p. 139, paragraphs III and IV, and *Bemerkung zu meinem zweitem Beitrag zur Theorie der Punktmengen*, Mathematische Annalen, vol. 65 (1908), pp. 431–432.

element in G distinct from the elements in the collection $\sum_{j=1}^{i-1} g_j$ which contains points in common with h_1, and let h_i be an arc whose end points are A and B, and which contains no points in common with g_i. Let F_0 be a closed point set which contains no points in common with $\overline{G}^*$. If i is a positive integer, it follows by part (2) of the hypothesis that if g_i^* is a maximal connected subset of $\bar{g}_i$, then it is a maximal connected subset of $\overline{G}^*$. Hence, by Theorem 2, there exists for each positive integer i a finite collection $F_i = f_{i1} + f_{i2} + \cdots + f_{in_i}$ of multiply connected Jordan domains covering g_i and such that (1) the upper distance of F_i from g_i is less than one half the smallest of the three positive numbers $1/i$, $d(g_i, G^* - g_i)$, and $d(g_i, \sum_{j=1}^{i-1} F_j)$; (2) the distance between any two distinct elements of F_i is a positive number; (3) the boundary, B_i, of F_i contains no point of $\overline{G}^*$. Let E be the set of all points of the arc h_1 belonging to $S - \sum_{i=1}^{\infty} \sum_{j=1}^{n_i} f_{ij}$; let C_i be the set of all points common to F_i and the arc h_i; and let $\overline{K} = E + \sum_{i=1}^{\infty} (B_i + C_i)$. It may readily be seen that $\overline{K}$ is a closed, bounded point set containing no point of G^*. Let K_1 be the maximal connected subset of $\overline{K}$ containing the point A. If K_1 does not contain B there exists, by Theorem 1, a simple closed curve J, containing no point of $\overline{K}$, such that one of the complementary domains of J contains A and the other contains B. Hence, J contains no points in common with the point set $\sum_{i=1}^{\infty} B_i$ and thus either (1) must be a subset of some domain f_{ij} of the collection $\sum_{i=1}^{\infty} \sum_{j=1}^{n_i} f_{ij}$ or (2) must be entirely without all the domains of this collection. The first case is impossible; for the arc h_i contains the points A and B, and thus contains a point x_i on J; x_i belongs to F_i, and hence to C_i and $\overline{K}$, thus contradicting the definition of J. Similarly, in the second case, J must contain a point of h_1 which is also a point of E and of $\overline{K}$. Thus, the supposition that K_1 does not contain B leads to a contradiction.

Hence, K_1 is a closed, connected, bounded point set containing A and B, but containing no point of G^*. It may readily be shown with the help of a theorem† due to R. L. Moore and a theorem‡ of Sierpiński's that K_1 is connected *im kleinen* at all of its points. It follows that K_1 is a continuous curve and that there exists within K_1 a simple continuous arc whose end points are A and B.§

---

† Cf. Moore, *Report on continuous curves from the viewpoint of analysis situs*, Bulletin of the American Mathematical Society, vol. 29 (1923), pp. 296–297, §3.

‡ Cf. W. Sierpiński, *Un théorème sur les continus*, Tôhoku Mathematical Journal, vol. 13 (1918), pp. 300–303.

§ Cf. R. L. Moore, *A theorem concerning continuous curves*, Bulletin of the American Mathematical Society, vol. 23 (1917), pp. 233–236; S. Mazurkiewicz, *Sur les lignes de Jordan*, Fundamenta Mathematicae, vol. 1 (1920), pp. 196–209; and R. Tietze, *Ueber stetige Kurven, Jordansche Kurvenbogen, und geschlossene Jordansche Kurven*, Mathematische Zeitschrift, vol. 5 (1919), pp. 284–291.

THEOREM 5. *If M is a closed and bounded point set, G is the aggregate of g's, where the symbol g represents a maximal connected subset of M, K is a bounded continuum containing at least one point in common with each element of G, and for each g, h_g is the point set $g \cdot \overline{(K-K\cdot M)}$ and H_g is a continuum containing h_g, then if $\overline{(K-K\cdot M+[H_g])}$ is bounded, it is a continuum.*

The theorem follows from Theorem 1 if M is connected. We shall suppose that G contains more than one element. Let g be a definite element of G, and e be a positive number which is less than $u(K, g)$. It can be shown with the help of Theorem 1 that there exist points of $K-K\cdot M$ within a distance of e from g. It follows that h_g is non-vacuous.

Let $N=\overline{(K-K\cdot M+[H_g])}$. If N is not connected, there exists, by Theorem 1, a simple closed curve J having no points in common with N and such that its interior D_1 and its exterior D_2 both contain points in common with N. If a point set h_i contains points in one of these domains, evidently H_i, and therefore h_i, is a subset of that domain. Let M_1 and M_2 respectively be the sums of all those elements g of G for which the corresponding H_g's are subsets of D_1 and D_2 respectively. The point set M_1 is closed. For, let z be a limit point of M_1 which does not belong to M_1. Since the elements of G are closed, there must exist an infinite sub-collection $\overline{G}_1$ of elements of G, such that (a) the sum of the elements of $\overline{G}_1$ is a subset of M_1, (b) every infinite sub-collection of $\overline{G}_1$ has a limiting set containing z, and (c) no element of $\overline{G}_1$ contains z. The limiting set F of G_1 is closed and connected.† It follows that F is a subset of an element of G. Let $\overline{h}_1$ be the collection of the h_i's corresponding to the elements of $\overline{G}_1$. We have shown that this collection is non-vacuous. The limiting set of h_1 is evidently a subset of F and of D_1. But this limiting set is also a subset of the h of that element of G which contains F, since every point in it is a limit point of $K-K\cdot M$. Hence F is a subset of an element of $\overline{G}_1$. It follows that M_1 and M_2 are closed and mutually separated.

Let K_1 and K_2 be those subsets of $(K-K\cdot M)$ which are subsets of D_1 and D_2 respectively. Since an "h" corresponding to an element of $\overline{G}_1$ is a subset of D_1, K_2 can have no point or limit point in M_1. By definition M_1 can have no limit point in K_2. It may readily be seen that the point sets $\overline{(M_1+K_1)}$ and $\overline{(M_2+K_2)}$ are mutually separated. But this contradicts the fact that $M+K=\overline{(M_1+K_1)}+\overline{(M_2+K_2)}$ is connected. Hence, the supposition that N is not connected has led to a contradiction.

DEFINITIONS. If K, M_1, and M_2 are point sets, K has no points in common with M_1+M_2 and no connected subset of $S-K$ contains points of both M_1

† Cf. Janiszewski, loc. cit.

and M_2, then we say that M_1 *and* M_2 *are separated by* K *or that* K *separates* M_1 *from* M_2. If K and M are mutually exclusive point sets and M is not a subset of any connected subset of $S-K$, then we say that K *separates* M. If K, H, and T are point sets, H and K are mutually exclusive, and for every positive number e and every point P of T there exists a positive number d_{eP} such that any two points of K whose distance from P is less than d_{eP} can be joined by a connected subset h_{eP} of $S-H$, whose upper distance from P is less than e, then we say that K *is not separated by* H *near* T.† If K and T are point sets, *then* K *is said to be connected near* T, provided that for every point P of T and every positive number e there exists a positive number d_{eP}, such that any two points of K whose distances from P are each less than d_{eP} are subsets of a connected point set h_{eP} which is a subset of K and which is at an upper distance less than e from P. If, in the preceding two definitions, it be specified that the connected point set h_{eP} be a continuum, then we say that K *is strongly not separated by* H *near* T, and K *is strongly connected near* T, respectively. If there exists for each positive number e a positive number d_{eP}, which, for the case of the definitions given in the preceding sentences, is independent of P, then we say respectively that K *is uniformly not separated by* H *near* T, and K *is connected near* T *uniformly*. If the point set T is closed and bounded, then, by an argument similar to that used in proving that a function which is continuous over a closed and bounded point set is uniformly continuous over that point set, it may be shown that if K is connected near T, or is not separated by H near T, then it has these properties uniformly over T. If H is closed and K is not separated by H near T, then K is strongly not separated by H near T; but this conclusion does not follow if H is not closed, as the following example will show. For $0 \leqq x \leqq 1$ let A_x, B_x, C_x, D_x, and E_x be the points with coördinates $(0, x)$, $(1, x)$, $(x, 0)$, $(x, 1)$, and $(x, 2)$ respectively; and let $K = A_0B_0 + C_0D_0E_0 + \sum_{i=1}^{\infty} C_{1/i}\ D_{1/i}\ E_{1/i}$, and let $H = S-(K-A_0)$. Then $K-A_0$ is not separated by H near A_0, but it is not true that $K-A_0$ is strongly not separated by H near A_0. Furthermore, $K-A_0$ is connected near A_0, but it is not true that $K-A_0$ is strongly connected near A_0.

If K is connected near T, and $T=K$, then K is everywhere connected im kleinen. If K and H are mutually exclusive, and T is any subset of $S-\overline{H}$, then K is not separated by H near T. If K is connected near T, and K and H are mutually exclusive, then K is not separated by H near T. If K is not separated by H near T, then neither is any subset of K.

† Zarankiewicz, *Sur les coupures locales faites par les continus*, Bulletin de l'Académie Polonaise des Sciences et des Lettres, Classe des Sciences Mathématiques et Naturelles, Série A: Sciences Mathématiques, 1927, p. 194, defines a concept with some similarity to ours but less general.

THEOREM 6. *If K is a bounded continuum, T is a closed subset of the boundary of $S-K$, T does not separate $K-T$, and $S-K$ is connected near T, then $K-T$ is strongly connected.*

If the theorem is not true, there must exist a pair of points x and y of $K-T$, such that any closed subset of $K-T$ containing x and y is not connected. Since T is closed, it can be proved, with the help of Theorems 10 and 15 of the paper F. A., that there exists a simple continuous arc xty containing no point of T. Let $0<e<\frac{1}{2}d(xty, T)$. Since T is closed, and $S-K$ is connected near T, $S-K$ has this property uniformly over T; and there exists a positive number d less than e, such that if P is a point of T, and x_1 and x_2 are a pair of points of $S-K$ whose distances from P are each less than d, then x_1 and x_2 belong to a connected subset of $S-K$ whose upper distance from P is less than e.

By Theorem 2 the point set T can be covered by a finite collection $H=h_1+h_2+h_3+\cdots+h_k$ of multiply connected Jordan domains, such that if m and n are distinct positive integers then (1) T contains no point of the boundary of h_m, (2) no point of h_m is at a distance from $T\cdot h_m$ greater than $d/10$, (3) the distance between h_m and h_n is positive. Let K^* be the set of all points belonging to K which are not covered by this collection of domains. The point set K^* is closed. Let X be the maximal connected subset of K^* containing the point x. Clearly X does not contain y. By Theorem 1 there exists a simple closed curve J_1 which separates x from y, and contains no point of K^*. Suppose that the interior, D_1, of this curve, contains y, and that its exterior, E_1, contains x. In the order xty on the arc xty let v be the first point common to K and D_1; in the order vx on the interval vx of this arc let z be the first point common to J_1 and xty, and w be the first point common to this interval and the product of K and E_1. Let z_1 be a point in the order wz_1z such that the interval wz_1 of wzv contains no point of J_1. It is easily shown that there exists a simple closed curve $z_1q_1zqz_1$ which has in common with the arc xty the two points z and z_1, encloses the segment z_1z of xty, but encloses or contains no other points of xty or of K. Let D be the complementary domain containing v of the sum of the simple closed curves J_1 and $z_1q_1zqz_1$. Its boundary J is a simple closed curve.† Then x is in the exterior

† J. Pál in *Om Planeus Topologi*, Matematisk Tidskrift B, 1923, pp. 66–72 (Festskrift til J. Hjelmslev), states the following theorem: *If the (simple, closed, plane) Jordan curves a and b have at least two points in common, then every complementary domain of their sum is bounded by a simple closed curve.* This result, while not explicitly stated by R. L. Moore, is easily derived from theorems he has proved; cf. either F. A., Theorem 41, p. 155, or *On the Lie-Riemann-Helmholtz-Hilbert problem of the foundations of geometry*, American Journal of Mathematics, vol. 41, pp. 299–319, Theorem 26.

E of J, J contains in common with wzv only the point z, and contains no points of K^*. Also, the arc wzv contains in common with K only the points w and v.

Let K_i be the product of h_i' and K. Let z_2 and z_3 be two distinct points of J distinct from z. In the order zz_2z_3 on J let A be the first point common to J and K_1, and let B be the first point on J in the order zz_3z_2 common to J and K_1. Let AzB be the interval AzB of J. Since h_1' is connected, there exists a finite chain of points $x_1(=A), x_2, x_3, \cdots, x_m(=B)$ belonging to h_1', such that the distance between any two consecutive points in this sequence is less than $d/10$. Let T_i be the product of h_i and T. From the definition of h_1 it follows that, for every value of i in the sequence 1, 2, 3, $\cdots$, m within a distance $d/10$ of x_i there exists a point y_i of T_1. Furthermore, since T_1 belongs to the boundary of $S-K$, there exists, within a distance $3d/10$ of y_i, a point t_i belonging to $S-K$. It is possible to select t_1 and t_m in such a way that t_1 belongs to the segment Az of the arc AzB, and that t_m belongs to the segment zB of this arc; for, since the curve J contains no points in common with K^*, the points A and B are limit points of the product of AzB and $S-K$.

For $i=1, 2, 3, \cdots, m-1$, the points t_i and t_{i+1} are each at a distance less than d from the point y_i of T_1. There exists a connected subset (t_i, t_{i+1}) of $S-K$, whose upper distance from y_i is not greater than e. Then $F=\sum_{i=1}^{m-1}(t_i, t_{i+1})$ is a connected subset of $S-K$ and its distance from the arc vzw is greater than e. It can be proved with the help of Theorems 10 and 15, F. A., that there exists a simple continuous arc $t_1\bar{t}t_m$ which is a subset of $S-K$ and contains no point whose distance from vzw is less than $\frac{1}{2}e$. The sum of the interval t_1zt_m of J and the arc $t_1\bar{t}t_m$ has as a subset a simple closed curve C_1 which (1) has no point in common with K^*+K_1, (2) has an interval E_1zF_1 in common with J, and (3) has in common with the arc vzw the point z and this point only. By a similar argument, this time using the curve C_1 instead of J, we can show the existence of a simple closed curve C_2 which has no point in common with $K^*+K_1+K_2$ and which has the properties (2) and (3) above. If we continue this process we get after k steps a simple closed curve C_k which (1) contains no point in common with $K=K^*+\sum_{i=1}^{k}K_i$, (2) has an interval E_kzF_k† in common with J, (3) has in common with the arc vzw the point z and the point z only. Since the arc vzw intersects‡ the arc E_kzF_k at z, v is in one complementary domain of C_k and w is in the other. Since C_k has no point in common with K, and K is connected, we have a contradiction. Hence, $K-T$ is strongly connected.

† Our notation implies that z is not an end point of E_kzF_k.

‡ We shall say that a point set A_1 *intersects* a point set A_2 at a point P provided that $A_1-A_1\cdot A_2$ is separated by A_2 near P.

THEOREM 7. *If T is a closed subset of a bounded continuum K, $K-T$ is connected, and $K-T$ is strongly connected near T, then $K-T$ is strongly connected.*†

Let x be a point of $K-T$. Let X be the set of all points of $K-T$ which can be joined to x by a closed and connected subset of $K-T$. Let $Y=K-T-X$. If Y is vacuous, the theorem must be true. If Y is non-vacuous, there must exist a point z which belongs to one of the point sets X and Y, and is a limit point of the other.

Suppose, first, that z belongs to X. Then z is the sequential‡ limit point of a sequence of points $y_1, y_2, y_3, \cdots$ belonging to Y. For each integer i let $h(y_i)$ be the set of all points in $K-T$ which can be joined to y_i by a continuum which is a subset of $K-T$. The point set $h(y_i)$ must have at least one limit point in T.§

By a theorem‖ of Janiszewski's the limiting set L of $\sum_{i=1}^{\infty} h(y_i)$ is closed and connected. Since T is closed and bounded, $K-T$ is strongly connected near T, uniformly. Hence, if $e=\frac{1}{2}d(z, T)$, there exists a positive number d less than e, such that if P is a point of T, and w_1 and w_2 are two points of $K-T$, each of whose distances from P is less than d, then there exists a closed and connected subset of $K-T$ containing w_1 and w_2 whose upper distance from P is less than e. Let L_1 be the maximal connected subset containing z of the point set consisting of all points of L whose distance from T is not less than $\frac{1}{2}d$. By an argument similar to that used in the preceding paragraph it follows that L_1 is closed, connected, and contains a point v_1 whose distance from T is $\frac{1}{2}d$. As v_1 belongs to the limiting set of $\sum_{i=1}^{\infty} h(y_i)$, there must exist an integer j such that $h(y_j)$ contains a point v_2 whose distance from v_1 is less than $\frac{1}{2}d$. There exist in $K-T$ three continua, k_1, k_2, and k_3, such that k_1 contains v_1 and v_2, k_2 contains v_2 and y_j, and k_3 contains

---

† The conclusions of this theorem and of Theorem 9 do not follow if the word "strongly" be omitted from the statement "$K-T$ is *strongly* connected near T," as the following example will show. Let K be the point set K mentioned in the paragraph preceding the statement of Theorem 6, and let T be the point A_0 there mentioned. Note, however, that this point set satisfies the hypothesis of Theorem 8.

The condition that $K-T$ be connected near T is not necessary for the conclusion of Theorem 7. Consider the set K mentioned above, but let T be the point E_0. Then $K-T$ is strongly connected, but is not connected near T.

‡ The point P is the *sequential limit point* of a sequence $P_1, P_2, P_3, \cdots$ of points, provided that every domain containing P contains all except a finite number of the points of the given sequence. Cf. F. A., p. 134.

§ Cf. R. L. Wilder, *Concerning continuous curves*, Fundamenta Mathematicae, vol. 7, p. 372, lines 17–33.

‖ Cf. Janiszewski, loc. cit.

z and x. Then $k_1+k_2+k_3+L_1$ is a continuum which contains x and y_j, and is a subset of $K-T$. This contradicts the definition of y_j. Hence z must belong to Y.

Hence there exists a sequence $x_1, x_2, x_3, \cdots$ of points belonging to X having z as a sequential limit point. For each positive integer, i, there exists in X a continuum l_i containing both x_i and x. Let L_2 be the limiting set of $\sum_{i=1}^{\infty} l_i$. If L_2 contains no points of T, then there exists in $S-K$ a continuum L_2 containing x and z, contrary to the fact that z belongs to Y. If L_2 contains points in common with T, we get a contradiction precisely as in the preceding paragraph. Thus Y is vacuous, and $K-T$ is strongly connected.

THEOREM 8. *If T is a set of condensation of a bounded continuum K, and $K-T$ is connected near T, then $K-T$ is connected.*†

If $K-T$ is not connected it is the sum of two mutually separated, non-vacuous point sets M and N. Then, since K is a continuum, $\overline{M}$ contains a limit point z of $K-\overline{M}$. Every point of T belonging to $K-\overline{M}$ is a limit point of N. Since $K-\overline{M}$ is a subset of $N+T$, z is a point or a limit point of N; also, z is a point or limit point of M. It follows from the hypothesis that there exists a circle about z, the interior of this circle containing a point of M and a point of N, these two points being joined by a connected subset of $K-T$. Thus the supposition that $K-T$ is not connected has led to a contradiction.

THEOREM 9. *If T is a closed set of condensation of a bounded continuum K, and $K-T$ is strongly connected near T, then $K-T$ is strongly connected.*

This theorem is a consequence of Theorems 8 and 7.

THEOREM 10. *Given that P is a point on the boundary H of a domain D, and that there exists a circle $(\overline{K})$ with center at P, and an infinite collection $D_1, D_2, D_3, \cdots$ of mutually exclusive maximal sub-domains of the product of D and the interior of $(\overline{K})$, and that P belongs to the limiting set of $D_1, D_2, D_3, \cdots$. Then if (C) and (K) are two circles concentric with such a circle, (C) is within $(\overline{K})$, and (K) is within (C), there exists on (C) an interval AE and on (K) an interval BF such that* (1) *there exists on the interval AE, in the order indicated, an infinite sequence $A_1, A_2, A_3, \cdots$ of points having A as a sequential*

† This theorem was suggested and proved by Dr. H. M. Gehman during a discussion of the proof of my Theorem 9. A point set T is said to be *a set of condensation* of a point set K, if every point of T is a limit point of $K-K\cdot T$. If K is closed, K contains T. Janiszewski, loc. cit., p. 102, defines in an analogous manner a *continuum* of condensation.

limit point, and on the interval BF, in the order indicated, an infinite sequence $B_1, B_2, B_3, \cdots$ *of points having B as a sequential limit point; (2) for each positive integer i there exists a simple continuous arc* A_iB_i *which is a subset of* D_i, *and contains in common with (C) only the point* A_i *and in common with (K) only the point* B_i; *(3) if* A_iA_{i+1} *and* B_iB_{i+1} *are the intervals* A_iA_{i+1} *and* B_iB_{i+1}, *respectively, of EA and FB, respectively,* J_i *is the simple closed curve* $A_iA_{i+1}B_{i+1}B_iA_i$, *and* E_i *is the interior of this curve, then* E_i *is a subset of the annular domain bounded by (C) and (K) and if* $i \neq j$, *then* E_i *and* E_j *have no points in common; (4) for each i there exists a sub-continuum* H_i *of H, such that* H_i *is a subset of* E_i', *and contains at least one point in common with* A_iA_{i+1} *and at least one point in common with* B_iB_{i+1}; *(5) the collections* $\sum_{i=1}^{\infty} H_i$ *and* $\sum_{i=1}^{\infty} A_iB_i$ *have a common sequential limiting set L, which is a sub-continuum of H; (6) no two elements in the collection* $\sum_{i=1}^{\infty} A_iB_i$ *can be joined by a connected subset, which is entirely within (K), of the domain D; (7) if x and y are points of two distinct elements of the collection* $L+\sum_{i=1}^{\infty} H_i$, *then there exists no connected subset of* $S-D$ *lying wholly within the annular domain bounded by (C) and (K), and containing both x and y.*

We leave the proof of this theorem to the reader; we refer him in particular to Theorem 1, and to an argument used by R. L. Wilder† in another connection.

THEOREM 11. *If P is a point on the boundary of a domain D, then a necessary and sufficient condition that P be accessible‡ from D is that there exist a subset K of D which is connected near P and has P as a limit point.*

Obviously the condition is necessary. We will proceed to show that it is sufficient. It follows from the hypothesis that there exists a sequence of points $x_1, x_2, x_3, \cdots$ belonging to K such that P is the sequential limit point of this sequence, and such that for each positive integer i there exists a connected subset of K containing both x_i and x_{i+1}, the upper distance of this connected subset from P being less than $1/i$. It follows by Theorems 10 and 15, F. A., that there exists a simple continuous arc x_ix_{i+1}, whose upper distance from P is less than $2/i$, and which is a subset of D. It follows by an

† Loc. cit., pp. 343–346.

‡ The point P on the boundary of a domain D is said to be *accessible* from that domain provided that for every point x of D there exists a simple continuous arc xP, whose end points are x and P, and which, except for the point P, is a subset of D. If for any simple continuous arc AB whatever, such that the segment AB is a subset of D, the point P is accessible from every maximal domain of $D-D\cdot AB$, which has P on its boundary, then P is said to be *accessible from all sides from D*. Cf. Schoenflies, *Entwickelung*, loc. cit., p. 176.

argument given by Wilder† and Theorem 15 F. A., that P is accessible from D.

THEOREM 12. *If B is the boundary of a domain D, P is a point of B, and D is not connected near P, then either* (1) *the hypothesis of Theorem* 10 *is satisfied at P with respect to the domain D and its boundary, or* (2) *the point set $B-P$ is not connected, and there exists a simple closed curve J which contains P, separates $B-P$, and, except for the point P, is a subset of D.*

Suppose that the hypothesis of Theorem 10 is not satisfied. Then if M is any circle about P, there exists a circle N^* with the same center such that there exist at most a finite number of maximal sub-domains of the product of D and the interior of M containing points within N.* Hence at least one of these domains contains P on its boundary. Let H_1 be such a domain with respect to a definite circle J_1 having its center at P, and $H_0=D$. It is impossible for H_1 and a circle Q within J_1 and concentric with it to satisfy the hypothesis of Theorem 10 at P; for, if they did, D and Q would do the same, and this involves a contradiction of our hypothesis. As above, it follows that the product of H_1 and the interior of Q has at least one sub-domain having P on its boundary. By a continuation of this process it is possible to select a sequence $J_1, J_2, J_3, \cdots$ of circles with centers at P and with radii approaching zero monotonically as n approaches infinity, and such that for each n there exists a maximal sub-domain H_n of the product of H_{n-1} and the interior of J_n having P on its boundary. Let y_n be a point of H_n such that if $i \neq j$ then $y_i \neq y_j$. There exists in H_n a simple continuous arc‡ joining y_n and y_{n+1}. It follows by Theorem 11 that there exists a simple continuous arc $y_1 tP$ which has end points y_1 and P, except for P is a subset of D, and for every n has in common with H_n a segment having P as an end point.

Consider the following proposition: *There exist two integers i and j, i a positive integer and j a positive integer or zero, such that H_j and the interior of J_i have in common at least two maximal sub-domains each of which has P on its boundary.* Suppose the proposition is false, and let N be a definite value of j and e be a positive number. Let $m>N$ be so chosen that $u(H_m, P)<e$. If H_N-H_m had P as a limit point, it would follow from the facts established at the beginning of the proof that H_N-H_m and the interior of J_m would have in common a maximal sub-domain having P as a boundary point, and thus that H_N and the interior of J_m would contain two such domains; this contradicts the assumption that the proposition stated at the beginning of this

† Loc. cit., pp. 343–344, Case 1.

‡ Cf. F. A., Theorem 15.

paragraph is false. It follows that there exists a circle with center at P and containing no points of $H_N - H_m$. Since H_m is a connected subset of H_N and its upper distance from P is less than e, where e is independent of N, it follows that H_N is connected near P. But this involves a contradiction of our hypothesis that $H_0 = D$ is not connected near P.

It follows that there exists an arc y_1wP of which y_1 and P are the end points such that $y_1wP - P \subset D$, the product of y_1tP and y_1wP is their end points, and that for n sufficiently large y_1wP contains no point of H_n. Let N be such an integer, and J be the sum of y_1tP and y_1wP. Then J is a simple closed curve. There exists a Jordan domain F_N which is a common subset of the interiors of J_N and J, and whose boundary is a subset of $J_N + J$ and contains P. It is easily seen that the boundary of F_N contains points in common with each of the segments y_1tP and y_1wP of J. Hence, since F_N is connected and is not a subset of D, it must contain a boundary point of D. Similarly, the exterior of J contains points of B. The truth of the theorem is thus established.

THEOREM 13. *If a point on the boundary of a domain is not accessible from all sides from that domain, then the hypothesis of Theorem* 10 *is satisfied at that point.*

Given a domain D_1, whose boundary H_1 contains a point P which is not accessible from all sides from D_1. Then there must exist a pair of points A and B belonging to $\overline{D}_1$, a simple continuous arc AXB, which, except for A and B, is a subset of D_1, and a sub-domain D of D_1 whose boundary H contains P, and is a subset of $H_1 + AXB$, the point P being not accessible from D. It is easily seen that all of the segment AXB belongs to H, and that if P is either A or B then P is accessible from D. Suppose that P is not a point of AXB. Then there exists a circle K_1 enclosing P but no point of AXB. It can be shown, with the help of Theorems 11 and 12, that there exists within K_1 a circle $(\overline{K})$ which has with reference to P, H, and D the properties mentioned in Theorem 10. It is easily seen that these properties hold also with reference to P, H_1, and D_1.

THEOREM 14. *In order that a bounded domain be connected near its boundary it is necessary and sufficient that it be uniformly connected im kleinen.*

It is easily seen that if a domain is uniformly connected im kleinen, then it is connected near its boundary. Let D be a bounded domain which is connected near its boundary, B. If D is not uniformly connected im kleinen, there must exist a positive number e and an infinite sequence of pairs of points, $x_1, y_1, x_2, y_2, x_3, y_3, \cdots, x_n, y_n, \cdots$, such that if n is any positive

integer, then $d(x_n, y_n) < 1/n$, but x_n and y_n cannot be joined by a connected subset of D, of diameter less than e. The set of x's has a limiting set containing at least one point, P. Evidently P belongs to B. There exists a number d_{eP} such that if x and y are both points of D at a distance less than d_{eP} from P, then x and y both belong to a connected subset of D, whose upper distance from P is less than $e/3$. Also, there exists an integer, n, such that the distances of x_n and y_n from P are each less than d_{eP}. Thus, the supposition that D is not uniformly connected im kleinen has led to a contradiction.

THEOREM 15. *In order that a domain be a Jordan domain it is necessary and sufficient that it be simply connected,† bounded, and connected near its boundary.*

The theorem follows from Theorem 14 and a theorem by R. L. Moore.‡

THEOREM 16. *If K is a bounded continuum, D is a complementary domain of K, H is a bounded subset of D, and $\overline{H} = H + T$, where T is a totally disconnected subset of K, then in order that there exist a simple closed curve, J, containing T, and separating $K - T$ from H, and such that $J - T$ is a subset of D, it is necessary and sufficient that K should not separate H near T.*

We shall first prove the sufficiency of the condition. Consider the case where $\overline{H}$ is a continuum, and H is connected near T. Since T is closed and bounded, H is connected near T, uniformly over T. It follows, with the help of Theorem 3, that for each positive integer n there exist a pair of point sets C_n and E_n defined as follows: C_1 is a circle, whose interior E_1 contains $K+H$; for n greater than unity E_n consists of the sum of a finite collection of Jordan domains covering T such that (1) each domain of this collection contains at least one point of T, and its diameter is less than $1/n$; (2) the distance between any two domains in this collection is a positive number; (3) the boundary, C_n, of E_n contains no point of T; (4) each domain of E_n plus its boundary is a subset of exactly one domain of E_{n-1}; (5) if t_1 and t_2 are points of H in or on the boundary of a maximal domain D^* of E_n, then there exists within that domain of E_{n-1} containing D^* a connected subset t_1t_2 of H containing t_1 and t_2. If P is a point of T, let $D(n, P)$ be that maximal domain of E_n which contains P, and let $J(n, P)$ be the boundary of $D(n, P)$. Let $\bar{x}$ and $\bar{y}$ be definite points of $K - T$ and H respectively. Let m be an integer greater than 10 such that E_m' contains neither $\bar{x}$ nor $\bar{y}$. Let K_1^* be the sum

† A domain is said to be simply connected if its boundary is connected.

‡ Cf. R. L. Moore, *A characterization of Jordan regions by properties having no reference to their boundaries*, Proceedings of the National Academy of Sciences, vol. 4 (1918), pp. 364–370.

of all those complementary domains of $K_1 = K + \sum_{i=m}^{\infty} C_i$ which have no limit points in H. Let $K^* = (K_1^* + K) - T$.

No limit point of K^* belongs to H. For, suppose H contains such a limit point Q. Then there exists a Jordan domain q containing Q but having no point or limit point in common with K, and containing points of not more than one simple closed curve belonging to the collection $\sum_{i=1}^{\infty} C_i$. If Q belongs to no such curve we may suppose that q contains no point of K_1. Then Q is a limit point of K_1^*. If q is a subset of some domain which is a subset of K_1^*, this domain has a limit point in H, contrary to its definition. If q contains points of at least two such domains, it must contain a boundary point of one of these domains, contrary to the hypothesis that q contains no point of K_1. Hence Q must belong to some simple closed curve J_1 of some C_i of the collection $\sum_{i=1}^{\infty} C_i$. Suppose again that q contains no point of K or of any simple closed curve of some C_i of the collection $\sum_{i=1}^{\infty} C_i$, distinct from J_1. Then there exists within J_1 a Jordan region R_1 and without J_1 a region R_2 such that the boundaries of these regions have in common a segment $t_1 Q t_2$ of J_1, containing the point Q, and such that R_1 and R_2 are subsets of q. It follows that either R_1 or R_2 must contain points of K_1^*. But, by an argument used above, it can be shown that this is impossible. Since all limit points of H, except those belonging to T, belong to H, the argument just given shows that K^* contains no points or limit points of H. Hence H and K^* are mutually separated. It is easily seen that $\overline{K}^* = K^* + T$ is a continuum.

By Theorem 8, H is connected. Let D^* be that complementary domain of K^* which contains H, and let $K_2 = S - D^*$. It is easily seen that K_2 contains K^* and is connected.

If D^* is not connected near T, there exists a point P of T, and a circle K_3 about P such that within every circle concentric with K_3, but having a smaller radius, there exist a pair of points of D^* which cannot be joined by a connected subset, lying entirely within K_3, of D^*. Since H is connected near T, there exists within K_3 a circle K_4, concentric with K_3, such that any two points of H within K_4 can be joined by a connected subset of H, lying entirely within K_3. There exists an integer n greater than $m+1$ such that $J(n, P)$ is a subset of the interior of K_4. There exists within $D(n, P)$ a pair, t_1 and t_2, of points of D^* which cannot be joined by a connected subset of D^*, lying entirely within K_3. The points t_1 and t_2 are points or boundary points of two domains d_1 and d_2, respectively, which are complementary domains of the point set K_1, and which are therefore subsets of $D(n, P)$. From the definition of K_1^* it follows that there exist a pair of points w_1 and w_2 of H, which are points or limit points of d_1 and d_2 respectively. There

exists a connected subset w_1w_2 of H, containing w_1 and w_2, and lying entirely within K_3. The point set $t_1+d_1+w_1w_2+d_2+t_2$ is a connected subset of D^*, and lies entirely within K_3. Thus, the supposition that D^* is not connected near T has led to a contradiction.

By Theorem 6, K_2-T is connected. By a theorem of Moore's† there exists a simple closed curve J which separates H from K_2-T, contains T, and, except for T, is a subset of D.

We have thus established the conclusion for the case where $\overline{H}$ is a continuum and H is connected near T. We shall next show that if H is any point set whatever satisfying the hypothesis, then there exists a point set H^* having the properties we have so far assumed for H. Let E_n, C_n, $\bar{y}$, and m have the same significance as before, with the exception that the fifth part of the definition of C_n and E_n read as follows: (5′) If t_1 and t_2 are points of H in or on the boundary of a maximal domain D^* of E_n, then there exists within that domain of E_{n-1} containing D^* *a connected subset* t_1t_2 *of* D *containing* t_1 *and* t_2. By the Heine-Borel-Lebesgue Theorem there exists *within* D a closed point set, $\overline{A}_{m+2}$, consisting of a finite number of circles plus their interiors, such that all points of H without E_{m+2} belong to $\overline{A}_{m+2}$. There exist at most a finite number of maximal connected subsets of $\overline{A}_{m+2}$. There exists within D a finite collection, $\overline{F}_{m+2}$, of simple continuous arcs, such that $\overline{A}_{m+2}+\overline{F}_{m+2}$ is connected. There exists in D a finite collection, $\overline{G}_{m+2}$, of arcs, such that $\overline{A}_{m+2}+\overline{F}_{m+2}+\overline{G}_{m+2}$ is connected, and that every maximal domain of E_{m+2} contains a point common to H and to $\overline{G}_{m+2}$. There exists a sequence $n_1(=m+2)$, n_2, n_3, $\cdots$ of positive integers, such that for each positive integer i, $n_i+2<n_{i+1}$, and such that if d_i is a maximal domain of the collection E_{n_i}, then d_i contains a point of H which does not belong to any domain of the collection $E_{n_{i+1}}$. Let M_{n_i} be the set of all points which belong to some domain of the collection E_{n_i} or the boundary of such a domain, but belong to no domain of the collection $E_{n_{i+2}}$. The set of points common to M_{n_i} and H is closed and can be covered by a finite set of Jordan regions, each region containing a point of $H\cdot M_{n_i}$ such that if A_{n_i} denotes the sum of these regions plus their boundaries, then A_{n_i} is a subset of both D and $M_{n_{i-1}}+M_{n_i}+M_{n_{i+1}}$. There exists by part (5) of the definition of E_n and Theorems 10 and 15, F. A., a finite collection of simple continuous arcs, whose sum F_{n_i} is a subset of D, and contains no point without $E_{n_{i-1}}$, such that if d_i is a maximal domain of E_{n_i} and R_1 and R_2 are connected subsets of A_{n_i} containing points in common with d_i, then there exists a closed and connected subset of $A_{n_i}+F_{n_i}$

† See Theorem 2, *Concerning the separation of point sets by curves*, loc. cit.

containing R_1 and R_2, and lying entirely within that maximal domain of E_{n_i-1} which contains d_i. Let

$$H^* = \overline{A}_{m+2} + \overline{F}_{m+2} + \overline{G}_{m+2} + \sum_{i=1}^{\infty}(A_{n_i}+F_{n_i}).$$

Then H^* contains H, is a subset of D, is connected, and is connected near T; also, $\overline{H^*}=H^*+T$. Hence, the condition in our theorem is sufficient. It follows from Theorem 15 that the condition is also necessary.

THEOREM 17. *If K and H are bounded continua which have in common a totally disconnected point set T, then a necessary and sufficient condition that there exist a simple closed curve which separates $K-T$ from $H-T$ is that $H-T$ is not separated by K near T.*

By a method of argument analogous to that used in proving Theorem 8 it can be shown that $H-T$ is a subset of a complementary domain of D. The truth of the theorem follows from Theorem 16.

THEOREM 18. *If T is a closed, totally disconnected subset of a bounded continuum K, and $K-T=K_1+K_2$, where K_1 and K_2 are mutually separated point sets, x is a point of K_1 and y of K_2, then there exists a simple closed curve which separates x from y and contains in common with K only points of T which are limit points of K_1.*

Let $T_1=T\cdot\overline{K}_1$. Then $K-T_1=K_1+K_2+T-T_1$. It follows that T_1 is closed and totally disconnected and that K_1 and K_2+T-T_1 are mutually separated point sets. Let Fy be a simple continuous arc containing no point of K_1, and having end points y and F, where F is a point of $S-K$. Let C be a circle with center at F, whose interior contains no point of K, I be an inversion of the plane with respect to the circle C, and H^*, K^*, T^*, r^*, x^*, and y^* be the transforms under this inversion of K_2+T-T_1, $\overline{K}_1$, T_1, $Fy-F$, x, and y. By methods similar to those used in the proof of Theorem 2 of Moore's paper *Concerning the separation of point sets by curves*,† we can show the existence of a simple closed curve J^* which separates x^* from y^*, encloses x^*, and contains no points of $H^*+K^*-T^*$. It is to be noted that both our hypothesis and our conclusion are weaker than Moore's, and that we are not concerned with the latter part of his argument. The inverse under I of J^* will be a simple closed curve of the type specified in the conclusion of the theorem.

Theorems 19–24, following, are corollaries of Theorem 18.

† Loc. cit., pp. 470–471.

THEOREM 19. *The conclusion of the preceding theorem remains true if the condition that K be bounded is removed, and the curve J be allowed to be either a simple closed curve or an open curve.*

THEOREM 20. *If K is a closed point set which consists of a collection of unbounded continua, and $K=K_1+K_2$, where K_1 and K_2 are mutually separated, and x is a point of K_1 and y is a point of K_2, then there exists an open curve which separates x from y and contains no point of K.*

Theorems 19 and 20 may be proved by performing an inversion of the plane about some sufficiently small circle with center at a point P which does not belong to K, and applying Theorem 18.

THEOREM 21. *If K is a bounded continuum and P is a cut point† of K, then there exists a simple closed curve which separates $K-P$; if K is an unbounded continuum there exists either a simple closed curve or an open curve which separates $K-P$.*

THEOREM 22. *A cut point of a continuum is accessible from some complementary domain of that continuum.*

THEOREM 23.‡ *If D is a simply connected domain, B is the boundary of D, and P is a cut point of B, then there exists either* (1) *a simple closed curve J which separates $B-P$, and which, except for P, is a subset of D, or* (2) *an open curve C which separates $B-P$ such that one ray on C from P is, except for P, a subset of D. If B is bounded, then condition* (1) *of the conclusion holds.*

THEOREM 24. *A cut point of the boundary of a simply connected domain is accessible from that domain.*

---

† The point P is a *cut point* of a connected point set K, if $K-P$ is not connected.

‡ Cf. Theorem 12.

UNIVERSITY OF TEXAS,
AUSTIN, TEXAS

TETRADS OF RULED SURFACES*

BY

A. F. CARPENTER

The projective differential geometry of a configuration composed of three ruled surfaces whose generators are in correspondence in sets of three has been discussed in two recent papers by the author.† It is the purpose of the present paper to present briefly an analytic basis for the projective differential geometry of a configuration composed of four ruled surfaces whose generators correspond in sets of four, each set containing one line element from each surface. This correspondence is brought about by choice of a parameter common to all four surfaces.

For defining system we choose a set of eight ordinary first-order linear and homogeneous differential equations in eight dependent variables, together with four linear and homogeneous equations of order zero:

$$(1)\qquad \alpha_i' = \sum_{j=i}^{i+3} a_{ij}\alpha_j, \qquad \alpha_{i+1}' = \sum_{j=i}^{i+3} a_{i+1,j}\alpha_j \qquad (i = 1,3,5,7),$$

$$(2)\qquad \sum_{l=1}^{8} b_{kl}\alpha_l = 0 \qquad (k = 1,2,3,4),$$

where a_{ij}, b_{kl} are functions of the independent variable x, and where all values of i, j, k or l greater than 8 are understood to be replaced with their residues, mod 8.‡

We take, further,

$$(3)\qquad a_{i,i+2}a_{i+1,i+3} - a_{i,i+3}a_{i+1,i+2} \neq 0 \qquad (i = 1,3,5,7),$$

and

$$(4)\qquad \begin{aligned} D_{l,l+1,l+2,l+3} &\equiv \left| b_{1l}b_{2,l+1}b_{3,l+2}b_{4,l+3} \right| \neq 0, \\ D_{l,l+1,l+4,l+5} &\equiv \left| b_{1l}b_{2,l+1}b_{3,l+4}b_{4,l+5} \right| \neq 0 \qquad (l = 1,3,5,7). \end{aligned}$$

A fundamental system of simultaneous solutions,

$$\alpha_i^{(h)}, \quad \alpha_{i+1}^{(h)} \qquad (h = 1,2,3,4\,;\ i = 1,3,5,7),$$

---

* Presented to the Society, San Francisco Section, June 2, 1928; received by the editors February 8, 1929.

† *Triads of ruled surfaces*, these Transactions, vol. 29 (1927). *Complete systems of invariants and covariants for triads of ruled surfaces*, Tôhoku Mathematical Journal, vol. 30 (1929).

‡ Throughout the discussion all single subscripts greater than 8 and all digits greater than 8 in multiple subscripts are understood to be replaced with their residues, mod 8.

of equations (1) can be interpreted as determining four lines $l_{i,i+1}$ given by the four pairs of points

$$(\alpha_i^{(1)}, \alpha_i^{(2)}, \alpha_i^{(3)}, \alpha_i^{(4)}), \quad (\alpha_{i+1}^{(1)}, \alpha_{i+1}^{(2)}, \alpha_{i+1}^{(3)}, \alpha_{i+1}^{(4)}).$$

By virtue of conditions (4) no two of these lines intersect. As x varies, these lines trace four ruled surfaces which, due to conditions (3), are non-developables.

It is to be noted that we can recover from equations (1) and (2), for any one of the four surfaces, the pair of second-order differential equations in two dependent variables which is fundamental in the projective differential geometry of ruled surfaces as developed by Wilczynski. For the surface R_{12} generated by l_{12} we annex to the original twelve equations those obtained by differentiating the pair

$$\alpha_1' = \sum_{j=1}^{4} a_{1j}\alpha_j, \qquad \alpha_2' = \sum_{j=1}^{4} a_{2j}\alpha_j.$$

From the set of fourteen equations thus derived, the twelve variables α_i', α_i, a_{i+1}', α_{i+1} $(i=3, 5, 7)$ can be eliminated, leaving the required pair of second-order equations in α_1, α_2. Finally from (1) and (2) by simple algebraic processes we can obtain the defining system for any of the four triads of surfaces, or any of the six pairs of surfaces, of the tetrad.

Except for conditions (3) no restrictions need be placed on the functions a_{ij} although it will be convenient to insist that they be analytic. For the functions b_{kl}, however, conditions (4) are not the only restrictions. If the four equations (2) be once differentiated and the derivatives of the dependent variables be replaced with their values from equations (1), there result four new equations of type (2). These, together with (2), constitute a linear homogeneous system of eight equations in the eight variables α_j and consistency demands the vanishing of the determinant D of the coefficients. This determinant involves the a_{ij} together with the b_{kl} and their first derivatives. Any choice of the b_{kl} consistent with the vanishing of this determinant and with conditions (4) will be suitable for our purpose.

The fact that equations (2) must be satisfied by any eight independent particular solutions of equations (1) suggests that we choose the b_{kl} as solutions of the system of differential equations adjoint to system (1) in such a way that

$$(5)\quad b_{kl}' = -\sum_{j=l+6}^{l+9} a_{jl} b_{kj}, \qquad b_{k,l+1}' = \sum_{j=l+6}^{l+9} a_{j,l+1} b_{kj} \quad (l = 1,3,5,7\ ;\ k = 1,2,3,4).$$

Since conditions (5) are consistent with (4) and cause the determinant D to vanish, they will be assumed to hold throughout the remainder of the discussion.

Permissible transformations on the variables are of the types standard in the projective differential geometry of ruled surfaces, that on the independent variable being

$$\bar{x} = \xi(x), \tag{6}$$

and those on the dependent variables,

$$\bar{\alpha}_i = c_i\alpha_i + c_{i+1}\alpha_{i+1}, \quad \bar{\alpha}_{i+1} = d_i\alpha_i + d_{i+1}\alpha_{i+1} \qquad (i = 1,3,5,7), \tag{7}$$

the coefficients being functions of x. For all systems obtained from (1) and (2) by transformations (6) and (7), conditions (3), (4), (5) will hold.

From (2), (6) and (7) we find that the expressions

$$b_{ki}\alpha_i + b_{k,i+1}\alpha_{i+1} \qquad (i = 1,3,5,7\ ;\ k = 1,2,3,4) \tag{8}$$

are absolute covariants and that

$$D_{jk,i} \equiv b_{ji}b_{k,i+1} - b_{ki}b_{j,i+1} \quad (i = 1,3,5,7\ ;j = 1,2,3\ ;k = 2,3,4\ ;j < k) \tag{9}$$

are relative invariants, since

$$D_{jk,i} \equiv \Delta_{(i+1)/2}\bar{D}_{jk,i}, \tag{10}$$

where

$$\Delta_{(i+1)/2} = c_id_{i+1} - c_{i+1}d_i,$$

and since the b_{kl} are invariant under transformation of the independent variable.

We are now in a position to transform system (1), (2) into an equivalent system whose coefficients will be absolute invariants. We choose for the transformation of the dependent variables

$$\bar{\alpha}_i = b_{1i}\alpha_i + b_{1,i+1}\alpha_{i+1},\ \bar{\alpha}_{i+1} = b_{2i}\alpha_i + b_{2,i+1}\alpha_{i+1} \quad (i = 1,3,5,7). \tag{11}$$

By making use of equations (5) there results the system

$$\begin{aligned} \bar{\alpha}_i' &= -\theta_{12,i}\bar{\alpha}_i + \theta_{11,i}\bar{\alpha}_{i+1} + \theta_{12,i+2}\bar{\alpha}_{i+2} - \theta_{11,i+2}\bar{\alpha}_{i+3}, \\ \bar{\alpha}_{i+1}' &= -\theta_{22,i}\bar{\alpha}_i + \theta_{21,i}\bar{\alpha}_{i+1} + \theta_{22,i+2}\bar{\alpha}_{i+2} - \theta_{21,i+2}\bar{\alpha}_{i+3}, \end{aligned} \tag{12}$$

$$\begin{aligned} &\sum_i \bar{\alpha}_i = 0, \qquad \sum_i \bar{\alpha}_{i+1} = 0, \\ &\sum_i (D_{32,i}\bar{\alpha}_i - D_{31,i}\bar{\alpha}_{i+1})/D_{12,i} = 0, \quad \sum_i (D_{42,i}\bar{\alpha}_i - D_{41,i}\bar{\alpha}_{i+1})/D_{12,i} = 0 \\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (i = 1,3,5,7), \end{aligned} \tag{13}$$

where $D_{jk,i}$ are given by (9), and where

$$(14)\quad D_{12,i}\theta_{kl,i} = a_{i+6,i}b_{k,i+6}b_{l,i+1} - a_{i+6,i+1}b_{k,i+6}b_{li} + a_{i+7,i}b_{k,i+7}b_{l,i+1} - a_{i+7,i+1}b_{k,i+7}b_{li} \quad (k = 1,2;\ l = 1,2).$$

As a consequence of (10) the coefficients of equations (13) are absolute invariants. It can be shown without difficulty that the coefficients of equations (12) are relative invariants of weight one. The effect of a transformation of the independent variable $\bar{x} = \xi(x)$ is to divide all coefficients of differential equations of the first order by ξ'. If then we choose ξ so that $\xi' =$ any $\theta_{kl,i}$, *equations* (13) *will be replaced with a set whose coefficients are all absolute invariants.*

Consider now any four space curves whose points correspond in fours, each four consisting of one point from each curve. If each such quadruple has the property that the four points belonging to it are coplanar, then we shall speak of the four curves as a quadrangular set. The geometric significance of transformation (11) is now apparent. *When the system of defining differential equations is of the form* (12), (13), *the tetrad of ruled surfaces* $R_{i,i+1}$ *is referred to eight covariant curves and these curves constitute two quadrangular sets, each set containing one curve from each surface.*

System (12), (13) is not unique since one of the same form could have been obtained by making use of any of the six transformations

$$(15)\quad \bar{\alpha}_i = b_{ki}\alpha_i + b_{k,i+1}\alpha_{i+1},\quad \bar{\alpha}_{i+1} = b_{li}\alpha_i + b_{l,i+1}\alpha_{i+1} \qquad \begin{pmatrix} i = 1,3,5,7; \\ k = 1,2,3;\ \ l = 2,3,4;\ \ k < l \end{pmatrix}.$$

In equations (13) there appear as coefficients sixteen absolute invariants involving the b_{kl} only and in equations (12), sixteen relative invariants of weight one, giving in all thirty-one absolute invariants involving the a_{ij} and b_{kl}. The six systems of type (12), (13) thus involve one hundred and eighty-six absolute invariants. But of invariants involving the a_{ij} and b_{kl} only, the theory calls for but thirty-one functionally independent, and investigation discloses that those already obtained in (12), (13) are independent. *Hence these constitute a fundamental system in terms of which all others of this type can be expressed.*

A fruitful source of invariants and covariants in this theory is found in the consideration of that property possessed by four skew straight lines of having only two straight line intersectors. The four lines $l_{i,i+1}$ $(i = 1, 3, 5, 7)$ determine two such intersectors. The pairs of points determined on each of the lines $l_{i,i+1}$ by these two intersectors are given by quadratic covariants and these covariants involve, in addition to the α_i, α_{i+1}, the coefficients of equations (2) [or (13)] only.

Considering next the ordinary tangents to the surface $R_{i,i+1}$ at points of $l_{i,i+1}$, there will be two which intersect both of the lines $l_{i+2,i+3}$, $l_{i+4,i+5}$, two which intersect both of the lines $l_{i+2,i+3}$, $l_{i+6,i+7}$, and two which intersect both of the lines $l_{i+4,i+5}$, $l_{i+6,i+7}$. There are thus determined nine pairs of covariant points on each line $l_{i,i+1}$, three of the nine corresponding to tangents to $R_{i,i+1}$ and the remaining six determined by tangents to the other three surfaces. Each such pair of points is given by a quadratic covariant which, in addition to involving the coefficients of equations (2) [or (13)] will be linear in the coefficients of equations (1) [or (12)], that is, of weight one.

Again, of the ordinary tangents to $R_{i,i+1}$ at points of $l_{i,i+1}$ there will be two which are tangents to $R_{i+2,i+3}$ at points of $l_{i+2,i+3}$, two which are tangents to $R_{i+4,i+5}$ at points of $l_{i+4,i+5}$ and two which are tangents to $R_{i+6,i+7}$ at points of $l_{i+6,i+7}$.* There are thus determined three pairs of covariant points on each line $l_{i,i+1}$, each pair given by a covariant which in addition to involving the coefficients of equations (2) [or (13)] will be quadratic in the coefficients of equations (1) [or (12)], that is, of weight two.

Considering finally the asymptotic tangents to $R_{i,i+1}$ at points of $l_{i,i+1}$, there are two which intersect $l_{i+2,i+3}$, two which intersect $l_{i+4,i+5}$, and two which intersect $l_{i+6,i+7}$.† There are thus determined on each line $l_{i,i+1}$ six pairs of covariant points, three pairs corresponding to asymptotic tangents to $R_{i,i+1}$ and the other three determined by asymptotic tangents to the other three surfaces. Each such pair of covariant points is given by a covariant of weight three.

To complete the count we may if we wish include the pair of flecnode points on $l_{i,i+1}$ which will correspond to a covariant of weight four. There are thus associated with each of the four lines $l_{i,i+1}$ $(i=1, 3, 5, 7)$ twenty quadratic covariants. The coincidence of the two points in any pair will be conditioned upon the vanishing of the discriminant of the corresponding covariant. All such discriminants are invariants.

Of the sequence of n points given by the expression

$$\phi_{(i+1)/2} = \kappa_{(i+1)/2}\alpha_i + \lambda_{(i+1)/2}\alpha_{i+1} \qquad (i = 1,3,5, \cdots, 2n-1),$$

those for which $n=4m+1, m=0,1,2, \cdots$, lie on the line l_{12} and the remaining lie in sets of three on the lines l_{34}, l_{56}, l_{78}, one on each line in order. If each point is such that the intersector tangent‡ at that point passes through the

* Lane, *Ruled surfaces with generators in one-to-one correspondence*, these Transactions, vol. 25, pp. 290, 291.

† See reference above.

‡ Lane, *Ruled surfaces in correspondence.*

next point in order, the n points constitute an intersector sequence.* We will suppose that the last point of the sequence is given by an n of the form $4m+1$ so that of the n points $m+1$ of them, including the first and the last, lie on l_{12}. If all $m+1$ are distinct, the n points constitute an open sequence of order m; if the last coincides with the first but the remaining $m-1$ are distinct, the n points constitute a closed sequence of order m.

A study of the conditions under which intersector sequences of orders $m=1, 2, 3, \cdots$ are closed gives rise to invariants involving only the a_{ij} and of weights expressible as functions of m. Methods of obtaining invariants of this type are available in a previous paper.†

We have seen above how on each line $l_{i,i+1}$ there are, in addition to the pair of flecnode points, nineteen other pairs which, from the method of their determination, are analogous to the flecnode points. The totality of these points for all line elements of $R_{i,i+1}$ constitutes nineteen surface curves each in general of two branches. Analogues to the complex points and the complex curves of a ruled surface are equally easy to define.

The four lines $l_{i,i+1}$ determine four quadrics $Q_{i,i+1}$ $(i=1, 3, 5, 7)$, each line lying on three of the four quadrics. Any two lines on R_{12} together with the three lines l_{34}, l_{56}, l_{78} determine a linear complex. As the two lines on R_{12} approach coincidenee this complex approaches a limit C_{12}. There are thus determined four linear complexes $C_{i,i+1}$ $(i=1, 3, 5, 7)$, and each line $l_{i,i+1}$ belongs to all four complexes.

To each point of l_{12} there corresponds the tangent plane at that point to Q_{34} and this plane contains l_{12}. Similarly, to each point of l_{12} there corresponds its polar plane with respect to C_{12} and this plane also contains l_{12}. The tangent plane and the polar plane for the same point are in general distinct but for two points on l_{12} they coincide. This pair of points is given by a quadratic covariant. Since the three quadrics and the four complexes can be paired as above in twelve different ways, there are thus determined on each line $l_{i,i+1}$ twelve pairs of covariant points giving rise on each surface $R_{i,i+1}$ to twelve curves which are analogues of the complex curve. The possibilities for such analogues are by no means exhausted. But there is no need to pursue this matter further. Additional sources of invariants and covariants readily suggest themselves and the theory as here outlined can be developed in as great detail as desired.

---

* *Triads of ruled surfaces.*

† *Triads of ruled surfaces.*

UNIVERSITY OF WASHINGTON,
SEATTLE, WASH.

ON THE CONGRUENCES CONNECTED WITH CERTAIN MAGIC SQUARES*

BY

D. N. LEHMER

The familiar "uniform step" method of constructing magic squares with odd number of cells† depends on certain interesting congruences which seem not to have been studied. By means of them one is able to state in very simple form the necessary and sufficient conditions that a square constructed with a given "step" and a given "break-step" shall be "magic," "diabolic," or "symmetric" completely or in part.

A square containing the numbers $1, 2, 3, \cdots, n^2$ is said to be "magic" in the rows if the sum of the numbers in each row is the same. Since the sum of all the numbers in the square is $n^2(n^2+1)/2$ the sum in each row must be $n(n^2+1)/2$. Similarly for the columns. A "magic square" is magic in both rows and columns. If the numbers in the diagonal from the upper left hand corner to the lower right hand corner as well as the numbers in any two partial diagonals parallel to it which together make up n numbers is the same the square is said to be magic in the positive diagonals. If it is also magic in the negative diagonals the square is called "diabolic." If further the sum of the two numbers in any two cells symmetrically placed with respect to the center cell is the same the square is said to be symmetric. As an illustration of a square which is magic, diabolic and symmetric one may take

14	23	39	6	15	31	47
18	34	43	10	26	42	2
22	38	5	21	30	46	13
33	49	9	25	41	1	17 .
37	4	20	29	45	12	28
48	8	24	40	7	16	32
3	19	35	44	11	27	36

The above square has been constructed by the "uniform step" method which may be described as follows:

---

* Presented to the Society, San Francisco Section, April 7, 1923, and June 12, 1926; received by the editors in January, 1929.

† See McClintock, *On the most perfect form of magic squares*, American Journal of Mathematics, vol. 19 (1897), p. 99.

Denote the cell in the pth column counting from the left and in the qth row counting from the bottom by (p, q). In the "uniform step" method of constructing magic squares one starts with the number 1 in the cell (p, q). The number 2 is then placed in the cell $(p+\alpha, q+\beta)$, the number 3 in the cell $(p+2\alpha, q+2\beta)$, and so on, the numbers $p+k\alpha$ and $q+k\beta$ being reduced modulo n. It is clear that by this process the number $n+1$ will land in the cell $(p+n\alpha, q+n\beta)$ and this, modulo n, is the same as (p, q) which is the cell already occupied by 1. A "break-step" (a, b) is then introduced which puts the number $n+1$ into the cell $(p+a, q+b)$. The original step (α, β) is then resumed and the numbers $n+1, n+2, n+3, \cdots$, are entered in the cells $(p+a, q+b)$, $(p+a+\alpha, q+b+\beta)$, $(p+a+2\alpha, q+b+2\beta), \cdots$. On arriving at the number $2n+1$ another resort to the break-step is necessary to avoid the already occupied cell $(p+a, q+b)$, and the numbers $2n+1, 2n+2, 2n+3, \cdots$ are entered in the cells $(p+2a, q+2b)$, $(p+2a+\alpha, q+2b+\beta)$, $(p+2a+2\alpha, q+2b+2\beta), \cdots$. The square resulting from a continuation of this process is seen to depend entirely on the four numbers α, β, a, b, and may or may not be magic, or if magic may or may not be diabolic or symmetric. We proceed to develope the necessary and sufficient relations between these four numbers to decide in advance the kind of square they indicate.

The result of the above process may be summed up in the following fundamental congruences for determining the cell (A, B) in which a number x is to be entered:

$$A \equiv p + \alpha(x-1) + a\left[\frac{x-1}{n}\right], \quad B \equiv q + \beta(x-1) + b\left[\frac{x-1}{n}\right] \pmod{n},$$

where (α, β) is the "step" and (a, b) is the "break-step," (p, q) is the cell containing the number 1 and the symbol $[k]$ denotes the greatest integer in k. In the square given above for illustration $(\alpha, \beta) = (1, 2)$, $(a, b) = (3, 5)$, $(p, q) = (6, 4)$, and the position of any number x is given by the congruences

$$A \equiv 6 + (x-1) + 3\left[\frac{x-1}{7}\right], \quad B \equiv 4 + 2(x-1) + 5\left[\frac{x-1}{7}\right] \pmod{7}.$$

It is to be observed, in the first place, that the process may not fill all the cells of the square. In this connection we have the following theorem:

Given α, β, a, b, all prime to n, the necessary and sufficient condition that the uniform step process shall fill the square is that the determinant $\left|\begin{smallmatrix}\alpha & a \\ \beta & b\end{smallmatrix}\right|$ shall be prime to n.

For if two values x_1 and x_2 of x should give the same cell (A, B), then

$$\alpha(x_1 - x_2) + a\left\{\left[\frac{x_1 - 1}{n}\right] - \left[\frac{x_2 - 1}{n}\right]\right\} \equiv 0 \pmod{n},$$

$$\beta(x_1 - x_2) + b\left\{\left[\frac{x_1 - 1}{n}\right] - \left[\frac{x_2 - 1}{n}\right]\right\} \equiv 0 \pmod{n},$$

whence, since b and β are supposed prime to n,

$$(\alpha b - \beta a)(x_1 - x_2) \equiv 0 \pmod{n} \tag{1}$$

and also

$$(\alpha b - \beta a)\left\{\left[\frac{x_1 - 1}{n}\right] - \left[\frac{x_2 - 1}{n}\right]\right\} \equiv 0 \pmod{n}. \tag{2}$$

If now $(\alpha b-\beta a)$ is prime to n, congruence (2) shows that x_1 and x_2 lie between the same two multiples of n. For since the values of $[(n-1)/n]$ range between 0 and $n-1$ the congruence

$$\left[\frac{x_1 - 1}{n}\right] \equiv \left[\frac{x_2 - 1}{n}\right] \pmod{n}$$

is also an equality. But if x_1 and x_2 lie between the same two multiples of n, congruence (1) is also an equality giving $x_1=x_2$ so that two numbers can not determine the same cell unless they are the same number. If however n and $(\alpha b-\beta a)$ have a common divisor δ then we have

$$\left[\frac{x_1 - 1}{n}\right] \equiv \left[\frac{x_2 - 1}{n}\right] \left(\mathrm{mod}\ \frac{n}{\delta}\right)$$

and $x_1 \equiv x_2 \pmod{n/\delta}$. If then we put $x_2=x_1+(n/\delta)n$ this last congruence is satisfied. Put this value of x_2 in the other congruence and we have

$$\left[\frac{x_1 - 1}{n}\right] \equiv \left[\frac{x_1 - 1}{n} + \frac{n}{\delta}\right] \left(\mathrm{mod}\ \frac{n}{\delta}\right)$$

and since n/δ is an integer the congruence is clearly satisfied. Two values of x which differ by n^2/δ will then fall in the same cell by this rule for filling the square and the square will therefore not be filled. In the theorems which follow this determinant is supposed to be prime to n and the square is therefore filled.

THEOREM. *If α and a are prime to n the square is magic in the columns.*

We show first the important fact that the numbers in any column form a complete system of incongruent numbers. For if two entries x_1 and x_2 occupy cells in the same column we must have

$$p + \alpha(x_1 - 1) + a\left[\frac{x_1 - 1}{n}\right] \equiv p + \alpha(x_2 - 1) + a\left[\frac{x_2 - 1}{n}\right] \pmod{n},$$

whence

$$\alpha(x_1 - x_2) + a\left\{\left[\frac{x_1 - 1}{n}\right] - \left[\frac{x_2 - 1}{n}\right]\right\} \equiv 0 \pmod{n}.$$

If now $x_1 \equiv x_2$ then since a is prime to n

$$\left[\frac{x_1 - 1}{n}\right] \equiv \left[\frac{x_2 - 1}{n}\right]$$

and, as before, this last congruence is also an equality and the two x's lie between the same two multiples of n. The congruence $x_1 \equiv x_2$ is also an equality and $x_1 = x_2$ contrary to hypothesis.

It is also clear that no two numbers in the same column can lie between the same two multiples of n, otherwise by the above congruence the two values of x would be equal.

If now we write $x_i = k_i n + l_i$ where $0 \leqq k \leqq n-1$ and $0 < l \leqq n$, then for the x's in any column the k's run through the values $0, 1, 2, 3, \cdots, n-1$ and the l's through the values $1, 2, 3, \cdots, n$, so that

$$\sum x_i = n \sum k_i + \sum l_i = \frac{n^2(n-1)}{2} + \frac{n(n-1)}{2} = \frac{n(n^2+1)}{2},$$

and the square is thus seen to be magic in the columns. Similarly if b and β are prime to n the square is magic in the rows. We have then the following theorem:

THEOREM. *If a, b, α, β are all prime to n the square is magic.*

This condition is also necessary. To show this we prove first that if a is prime to n but α and n have the greatest common divisor δ, then the square is not magic in the columns. From the congruence obtained above for any two entries x_1 and x_2 in the same column

$$\alpha(x_1 - x_2) - a\left\{\left[\frac{x_1 - 1}{n}\right] - \left[\frac{x_2 - 1}{n}\right]\right\} \equiv 0 \pmod{n},$$

we can see that if k is the smallest value of $[(x-1)/n]$ in the column then all the others in that column are given by $k + \rho_i \delta$ where $\rho_i = 0, 1, 2, \cdots, n/\delta - 1$.

For if the congruence

$$\alpha(x-1) \equiv A - p - ak \pmod{n}$$

has solutions, so will also the congruence

$$\alpha(x-1) \equiv A - p - a(k+\rho_i\delta) \pmod{n}.$$

Moreover for each value of ρ_i there will be just δ solutions. Writing as above $x_i = k_i n + l_i$ it is clear that the values of k_i fall into δ sets of n/δ numbers $k+\rho_i\delta$, so that

$$\sum k_i = \delta \sum k + \rho_i\delta = \delta\left\{k\frac{n}{\delta} + \frac{\delta}{2}\left(\frac{n}{\delta}\right)\left(\frac{n}{\delta} - 1\right)\right\}$$

$$= kn + \frac{n}{2}(n-\delta).$$

As for the l_i, they form again a complete system of incongruent numbers. For if two were congruent then by the above reasoning the corresponding x's would be equal. The l's give again the series $1, 2, 3, \cdots, n$ in some order and $\sum l_i = n(n+1)/2$. We have then for the sum of the x's in the column

$$\sum x_i = \sum k_i n + l_i = kn^2 + \frac{n^2}{2}(n-\delta) + \frac{n(n+1)}{2}$$

$$= \frac{n}{2}(n^2+1) + n^2\left(k - \frac{\delta-1}{2}\right).$$

Now k, the smallest value of $[(x-1)/n]$ that will make $A-p-a[(x-1)/n]$ divisible by δ, is seen to depend on A and as A takes the values $1, 2, 3, \cdots, n$, k will run through the series of numbers $0, 1, 2, \cdots, \delta$ in some order, and two values of A which are congruent modulo δ will give the same k. We have then the following theorem:

THEOREM. *If a is prime to n but α and n have the greatest common divisor δ then the square is not magic in the columns but the sum in any column differs from $(n/2)(n^2+1)$ by some multiple of n^2 and column A and column $A+\delta m$ have the same sum.*

In the above formula if $\delta = 1$ then $k = 0$ and the sum of the x's in the column is $(n/2)(n^2+1)$ as it should be.

A similar theorem holds, of course, for the rows when b is prime to n while β and n have the greatest common divisor δ. As an illustrative example we take the square defined by $(\alpha, \beta) = (3, 4)$, $(a, b) = (2, 3)$, $(p, q) = (3, 4)$, $n = 15$:

168	111	54	222	165	93	36	204	147	90	18	186	129	72	15
125	68	11	179	107	50	218	161	104	32	200	143	86	29	182
82	25	193	121	64	7	175	118	46	214	157	100	43	196	139
39	217	150	78	21	189	132	75	3	171	114	57	225	153	96
221	164	92	35	203	146	89	17	185	128	71	14	167	110	53
178	106	49	217	160	103	31	199	142	85	28	181	124	67	10
135	63	6	174	117	60	213	156	99	42	210	138	81	24	192
77	20	188	131	74	2	170	113	56	224	152	95	38	206	149
34	202	145	88	16	184	127	70	13	166	109	52	220	163	91
216	159	102	45	198	141	84	27	195	123	66	9	177	120	48
173	116	59	212	155	98	41	209	137	80	23	191	134	62	5
130	73	1	169	112	55	223	151	94	37	205	148	76	19	187
87	30	183	126	69	12	180	108	51	219	162	105	33	201	144
44	197	140	83	26	194	122	65	8	176	119	47	215	158	101
211	154	97	40	208	136	79	22	190	133	61	4	172	115	58

In accordance with the theorem the square is not magic in the columns, since the greatest common divisor of α and n is 3. In accordance with a theorem to be proved it is also not magic in the rows because the greatest common divisor of b and n is also 3. In the first column the smallest k is 2 corresponding to the three entries 39, 34, 44. The other k's are 5, corresponding to the three entries 82, 77, 87; and 8 corresponding to the entries 125, 135, 130; and 11 corresponding to the entries 168, 178, 173; and 14 corresponding to 221, 216, 211. The sum in the first column is 1920 as given by the formula for $k=2$, $\delta=3$ and $n=15$. In the second column $k=1$ and the sum is 1695 by the formula. In the third column $k=0$ and the sum is 1470. The sum is the same in the 1st, 4th, 7th, 10th, and 13th columns. Also in the 2d, 5th, 8th, 11th, and 14th. Also in the 3d, 6th, 9th, 12th, and 15th.

We show now that if α is prime to n but a and n have the greatest common divisor δ, then the square is not magic in the columns. Starting with the same congruence which holds between two entries in the same column,

$$\alpha(x_1 - x_2) + a\left\{\left[\frac{x_1 - 1}{n}\right] - \left[\frac{x_2 - 1}{n}\right]\right\} \equiv 0 \pmod{n},$$

we see that $x_1 \equiv x_2 \pmod{\delta}$ and if l is the smallest non-zero residue of the x's in the column, then all the l's in the column may be written $l+\rho_i\delta$ where $\rho_i = 0, 1, 2, \cdots, n/\delta-1$. We have then δ sets of numbers $l+\rho_i\delta$ and the sum of the l's in the column is

$$\sum l_i = \delta \sum_{i=0}^{n/\delta-1} (l + \rho_i\delta) = nl + \frac{n(n-\delta)}{2}.$$

As before the k's form a complete series of incongruent numbers running through the values 0, 1, 2, $\cdots$, $n-1$. Their sum is therefore $n(n-1)/2$. We thus get for the sum of the numbers in the column

$$\sum x_i = \sum k_i n + l_i = \frac{n^2(n-1)}{2} + nl + \frac{n(n-\delta)}{2}$$

$$= \frac{n}{2}(n^2+1) + nl - \frac{n(n-\delta)}{2}.$$

As before it is clear that as the column A is changed the values of l run through the series of numbers 1, 2, 3, $\cdots$, δ and any two columns A and $A+\delta m$ will have the same l. It should be noted, however, that l appears in the formula as multiplied by n and not by n^2 so that the sums in the different columns will differ by multiples of n and not as in the preceding case by n^2. We may then state the following theorem:

THEOREM. *If α is prime to n but a and n have the greatest common divisor δ, then the square is not magic in the columns but the sum differs from $(n/2)(n^2+1)$ by some multiple of n and column A and column $A+\delta m$ have the same sum.*

A similar theorem holds of course for the rows when a is prime to n and b and n have the greatest common divisor δ. In the illustrative example it is seen that the top row has the sum 1710 which comes from the formula by putting $n=15$, $\delta=3$, $l=3$. The same sum appears in the 4th, 7th, etc. rows from the top. The second row from the top adds to 1695 and the third to 1680.

The preceding results may then be summed up in the following fundamental theorem:

THEOREM. *The necessary and sufficient condition that the square obtained from the congruences*

$$A \equiv p + \alpha(x-1) + a\left[\frac{x-1}{n}\right],$$

$$B \equiv q + \beta(x-1) + b\left[\frac{x-1}{n}\right] \pmod{n}$$

shall be magic is that the five numbers α, β, a, b and $(\alpha b - \beta a)$ shall be prime to n.

This theorem makes it clear why it is not possible to construct by the uniform step method a square with an even number of elements. In fact the five numbers indicated can not all be odd. It should also be noted that the unit cell (p, q) does not enter into the condition at all. We can therefore

denote the above square by $\left(\begin{smallmatrix}\alpha & a\\ \beta & b\end{smallmatrix}\right)$ with the understanding that the initial cell may be chosen arbitrarily.

Conditions for diabolism. If $A+B\equiv$constant it is easily seen that the cells (A, B) all lie in the same positive diagonal. Indeed, to keep the sum constant we must subtract from B anything which we add to A. Similarly if $A-B\equiv$constant the cells (A, B) all lie in the same negative diagonal. Now we have

$$A + B \equiv p + q + (\alpha + \beta)(x - 1) + (a + b)\left[\frac{x-1}{n}\right],$$

$$A - B \equiv p - q + (\alpha - \beta)(x - 1) + (a - b)\left[\frac{x-1}{n}\right] \pmod{n}.$$

If therefore the numbers $\alpha+\beta$, $\alpha-\beta$, $a+b$, $a-b$, and $(\alpha+\beta)(a-b)-(\alpha-\beta)\cdot(a+b)$ are prime to n these congruences indicate a magic square whose columns are the positive diagonals and whose rows are the negative diagonals of the given square. We have then the following theorem:

THEOREM. *The magic square* $\left(\begin{smallmatrix}\alpha & a\\ \beta & b\end{smallmatrix}\right)$ *is diabolic if and only if the numbers* $\alpha\pm\beta$, $a\pm b$ *are prime to* n.

(The condition that $(\alpha+\beta)(a-b)-(\alpha-\beta)(a+b)$ be prime to n is fulfilled automatically if $\alpha b-\beta a$ is prime to n.) A curious corollary of this theorem is that a diabolic square cannot be constructed by the uniform step process when n contains a factor 3. For if α and β are prime to 3, $\alpha+\beta$ and $\alpha-\beta$ cannot both be. This does not mean, however, that a diabolic square cannot be obtained by other methods when n is a multiple of 3.

Conditions for symmetry. In a symmetrical square the sum of the two entries x_1 and x_2 in any two cells (A, B) and $(1-A, 1-B)$ symmetrically situated with respect to the center is equal to n^2+1. This definition leads to the congruences

$$A \equiv p + \alpha(x_1 - 1) + a\left[\frac{x_1 - 1}{n}\right],$$

$$B \equiv q + \beta(x_1 - 1) + b\left[\frac{x_1 - 1}{n}\right],$$

$$1 - A \equiv p + \alpha(n^2 + 1 - x_1 - 1) + a\left[\frac{n^2 + 1 - x_1 - 1}{n}\right],$$

$$1 - B \equiv q + \beta(n^2 + 1 - x_1 - 1) + b\left[\frac{n^2 + 1 - x_1 - 1}{n}\right] \pmod{n}.$$

The third and fourth simplify to

$$1 - A \equiv p - \alpha x_1 + a\left[n - \frac{x_1}{n}\right],$$

$$1 - B \equiv q - \beta x_1 + b\left[n - \frac{x_1}{n}\right] \qquad (\text{mod } n).$$

From these we get, with the first and second,

$$1 \equiv 2p - \alpha + a\left\{\left[\frac{x_1 - 1}{n}\right] + \left[n - \frac{x_1}{n}\right]\right\},$$

$$1 \equiv 2q - \beta + b\left\{\left[\frac{x_1 - 1}{n}\right] + \left[n - \frac{x_1}{n}\right]\right\} \qquad (\text{mod } n).$$

But we can show that

$$\left[\frac{x_1 - 1}{n}\right] + \left[n - \frac{x_1}{n}\right] = n - 1.$$

For suppose $(x_1-1)/n=k+\rho$ where k is an integer and $0\leqq\rho\leqq(n-1)/n$. Then the above expression is $k+[(n-k)-(\rho+1/n)]$ and since $0<\rho+1/n\leqq 1$, this is $k+n-k-1$ or $n-1$. The above congruences then become

$$1 \equiv 2p - \alpha - a, \qquad 1 \equiv 2q - \beta - b \qquad (\text{mod } n),$$

and we have the following theorem:

THEOREM. *The magic square* $\left(\begin{smallmatrix}\alpha & a\\ \beta & b\end{smallmatrix}\right)$ *is symmetric if and only if* $2p\equiv\alpha+a+1$, $2q\equiv\beta+b+1$ (mod n).

The conditions for symmetry do not depend in any way upon the condition for a square to be magic, so that a square may be symmetric and not magic, as is exemplified in the above square of 15. The central square pairs with itself and must be $(n^2+1)/2$.

A more general type of diabolism arises by connecting the coördinates of the cells (A, B) by a linear relation $\lambda A+\mu B\equiv$ constant (mod n). Cells related in this way may be said to form a diagonal with a complementary diagonal given by $\lambda A-\mu B\equiv$ constant (mod n). With these diagonals we have the

THEOREM. *The square* $\left(\begin{smallmatrix}\alpha & a\\ \beta & b\end{smallmatrix}\right)$ *is diabolic with respect to the diagonals* $\lambda A\pm\mu B$ *if and only if the numbers* $\lambda\alpha\pm\mu\beta$, $\lambda a\pm\mu b$ *are prime to* n.

Enumeration of uniform step squares. By means of the fundamental congruences defining uniform step squares it is easy to see that the tota number of such for a given prime p is $p^2(p-1)^3(p-2)$. For we may choose

the unit cell in p^2 different ways. Also the numbers α, β, a may be chosen at will from the $p-1$ numbers less than p. For each such choice there are $p-2$ choices of b since the determinant $\alpha b-\beta a$ must be different from zero modulo p. The above result may be written $p^2\phi_1^3(p)\,(\phi_1(p)-1)$ where ϕ_1 is Euler's totient function.

For the case where the order of the square is a composite number n the enumeration must be given more detailed consideration. In the congruence $\alpha b-\beta a\equiv k \pmod{n}$, α, β and a are given relatively prime to n; then as k runs through a set of values prime to n we are to find the number of b's which are also prime to n. This problem is equivalent to the following:

Let λ be one of the numbers λ_n which are less than a given number n and prime to it. Find the number of numbers ϕ_2 in the set $\lambda+\lambda_n$ which are prime to n.

Consider the complete set of numbers $1, 2, 3, \cdots, n$. We erase first of all the n/p_1 multiples of p_1, where p_1 is a prime divisor of n. When λ is added to the set of numbers $1, 2, \cdots, n$ we get again a complete set of incongruent numbers modulo n from which again n/p_1 multiples of p_1 must be erased. For the prime p_1 we have therefore erased $2n/p_1$ numbers from the list of numbers $1, 2, \cdots, n$, leaving $n-2n/p_1$. Similarly for the other prime factors of n. But any multiple of p_1p_2 has by this process been erased twice. We must then add $4n/(p_1p_2)$, etc. By the well known principle of cross classification this gives at once the formula for ϕ_2:

$$\phi_2 = n - \sum\frac{2n}{p_1} + \sum\frac{4n}{p_1p_2} - \sum\frac{8n}{p_1p_2p_3} + \cdots,$$

which may be written in the form

$$\phi_2 = n\Pi\left(1-\frac{2}{p_i}\right)$$

or in the form

$$\phi_2 = \Pi p_i^{\alpha_i-1}(p_i-2).$$

The formula gives zero for an even number which is as it should be, because in this case the numbers less than n and prime to it are all odd and the sum of any two of them must be even and so not prime to n. As an illustration of the theorem take $n=45$. The set of numbers less than 45 and prime to it is

1, 2, 4, 7, 8, 11, 13, 14, 16, 17, 19, 22, 23, 26, 28, 29, 31, 32, 34, 37, 38, 41, 43, 44.

Add to these any one of the list, say 11, and erase all those that are not prime to 45, and the nine numbers

13, 19, 22, 28, 34, 37, 43, 49, 52

are all that remain. The formula gives

$$\phi_2(45) = 3(3-2)(5-2) = 9.$$

The above formula for ϕ_2 is the same as that found by V. Schemmel (Journal für die reine und angewandte Mathematik, 1869, pp. 191–192) for the number of pairs of consecutive numbers each less than n and both prime to n. It is seen by the above discussion that the numbers are not necessarily consecutive. Schemmel also gives the formula for the number of sets of k consecutive integers all of which are prime to n, which is

$$\phi_k = \Pi p_i^{\alpha_i - 1}(p_i - k).$$

This formula is the enumerative function for the following much more general problem:

Given a set of k linear forms a_1x+b_1, a_2x+b_2, a_3x+b_3, $\cdots$, a_kx+b_k, *where the a's are all prime to n and incongruent* mod n. *Let x run through a complete set of incongruent numbers. Find the number of values of x for which all the k forms will give numbers prime to n.*

To prove this formula we observe that as x runs through a complete set of incongruent numbers so also does $ax+b$ and that there are n/p_1 values of $ax+b$ which are multiples of p_1 where p_1 is any prime factor of n. Each one of these values of $ax+b$ indicates a value of x which is to be erased from the numbers 1, 2, 3, $\cdots$, n. The a's being all supposed incongruent modulo n we will thus have kn/p_1 numbers to erase from the set 1, $\cdots$, n. The argument is thus the same as in the special case considered above. Many theorems regarding the ordinary totient function have their analogues for the more general function. Thus evidently if m and n are relatively prime we have $\phi_k(m)\phi_k(n) = \phi_k(mn)$.

The value of $\phi(1)$ being assumed equal to unity, the theorem *The sum of totients of the divisors of n is equal to n* carries over only when n contains no square factor other than unity. But in this case we have the theorem

The sum of the ϕ_k *functions of the divisors of n is equal to* ϕ_{k-1} *function of n when and only when n has no square divisors other than unity.*

We are now able to write down the number of magic squares obtainable by the uniform step method for any number $n = \Pi p_i^{\alpha_i}$. The formula is

$$n^2\phi_1{}^3(n)\phi_2(n) = n^2[\Pi p_i^{\alpha_i - 1}(p_i - 1)]^3 \Pi p_i^{\alpha_i - 1}(p_i - 2).$$

For symmetric squares we have no choice of the unit square. The formula for symmetric squares is therefore this result divided by n^2, or $\phi_1{}^3(n)\phi_2(n)$.

For diabolic squares we may choose α and a at will from the $\phi_1(n)$ numbers less than n and prime to it. Then β must be chosen so that $\beta, \alpha+\beta$ and $\alpha-\beta$ shall be prime to n. This gives $\phi_3(n)$ choices for β. Finally b must be chosen so that the four numbers b, $a+b$, $a-b$, $\alpha b-\beta a$ shall be prime to n. This number is $\phi_4(n)$. The formula, therefore, for the total number of diabolic squares obtainable by the uniform step method is

$$\phi_1^2(n)\phi_3(n)\phi_4(n) = [\Pi p_i^{\alpha_i-1}(p_i-1)]^2\Pi p_i^{\alpha_i-1}(p_i-3)\Pi p_i^{\alpha_i-1}(p_i-4).$$

This last formula points out the fact which we have noted before, that no diabolic square is obtainable by the uniform step method when n is a multiple of 3. The following table gives the values of these functions for odd values of n up to $n=35$.

n	Magic squares	Symmetric squares	Diabolic squares
3	72	8	0
5	4800	192	32
7	52920	1080	432
9	52488	648	0
11	1089000	9000	5600
13	3212352	19008	12960
15	345600	1536	0
17	17756160	61440	46592
19	35790984	99144	77760
21	3810240	8640	0
23	118288632	223608	183920
25	75000000	12000	20000
27	38263752	52488	0
29	498464064	592704	509600
31	752463000	783000	680400
33	78408000	72000	0
35	254016000	207360	13824

Transformations of magic squares. Certain transformations on the numbers 1 to n^2 have been devised which serve to throw a magic square into another magic square (see McClintock, loc. cit.). The question has not been studied as to whether the resulting square is a uniform step square or not. We proceed to answer this question by finding the connection between these transformations and the constants α, β, a, b, which determine the square.

The numbers 1 to n^2 fall into n sets as follows:

$$
\begin{aligned}
K_0 &\quad (1, 2, 3, \cdots, n),\\
K_1 &\quad (n+1, n+2, \cdots, 2n),\\
K_2 &\quad (2n+1, 2n+2, \cdots, 3n),\\
&\quad \cdots\cdots\cdots\cdots\cdots\cdots\\
K_{n-1} &\quad (n(n-1)+1, n(n-1)+2, \cdots, n^2).
\end{aligned}
$$

Let now s be any substitution in the numbers 1 to n such as

$$s = \begin{pmatrix} 1, & 2, & 3, \cdots, & n \\ x_1, & x_2, & x_3, \cdots, & x_n \end{pmatrix}.$$

Perform now the same substitution on the numbers in each of the sets K. We get thus a transformation of the numbers 1 to n^2 which we denote by S. This transformation S, as we can show, will throw a uniform step square into a magic square which may or may not be a uniform step square. We have seen, in fact, that in a uniform step square each row (or column) contains one and only one number out of each of the above sets K, also that the numbers in any row (or column) form a complete series of incongruent numbers modulo n. Further, these two facts are sufficient to ensure that the square be magic. Now it is clear that neither of these conditions is disturbed by the transformation S, for the sets are not broken up, and also two entries in the same line can not be congruent to each other, for if $kn+x_1$ and $kn+x_2$ were congruent modulo n so would also be x_1 and x_2 which is not so. The new square is therefore magic.

Again, let T be any substitution on the sets K. Then again, if we start with a uniform step square the transformed square is still magic but not necessarily a uniform step square. For as before each set will furnish one and only one number in any row or column and those in any row or column will still furnish a complete set of incongruent numbers modulo n.

There are manifestly $(n!)^2$ transformed magic squares obtainable by S and T transformations from any given uniform step square. No two squares obtained from the same square by two different substitutions can be identical, for two different substitutions S must make a different rearrangement of the numbers 1 to n and so will give different squares. So also for two different transformations T. Moreover a transformation S can not produce the same square as a transformation T (except, of course, the identity transformation); for a transformation T can not rearrange the numbers in the separate sets.

We consider now the following question: *When will an S or a T transformation throw a uniform step square into another uniform step square?*

Let s be a substitution on the numbers 1 to n such as

$$s = \begin{pmatrix} x_1, & x_2, & x_3, & \cdots, & x_n \\ y_1, & y_2, & y_3, & \cdots, & y_n \end{pmatrix}$$

where the x's and y's are correlated by the congruence $y \equiv \lambda x + \rho \pmod{n}$ where λ is prime to n. Two values of y give $y_1 - y_2 \equiv \lambda(x_1 - x_2) \pmod{n}$ so that the difference between consecutive y's is λ times the difference between the corresponding x's. This is equivalent to saying that the coefficients which determine the step have both been multiplied by a constant. The "break-step" has, however, not been changed.

A transformation T similarly defined makes a corresponding change in the coefficients of the "break-step" leaving the "step" unaltered. It is not difficult to see also that, if the difference between the y's is to be constant for consecutive x's, a congruential relation of the sort indicated above must hold. We have then the

THEOREM. *If the transformation S or T is based on the substitution*

$$s = \begin{pmatrix} x_1, & x_2, & x_3, & \cdots, & x_n \\ y_1, & y_2, & y_3, & \cdots, & y_n \end{pmatrix},$$

then the necessary and sufficient condition that the resulting square be a uniform step square is that the x and y be related by the congruence $y \equiv \lambda x + \rho \pmod{n}$ where λ is prime to n.

The above result may be stated as follows:

Those uniform step squares which are obtainable from a given square $\binom{a\ a}{\beta\ b}$ by an S-transformation are of the form $\binom{k\alpha,\ a}{k\beta,\ b}$, and those obtainable by a T-transformation are of the form $\binom{a,\ k'a}{\beta,\ k'b}$.

The square $\binom{a\ a}{\beta\ b}$ and the square $\binom{k\alpha,\ k'a}{k\beta,\ k'b}$ may be called equivalent under H and V. Since the constants k and k' may take $\phi_1(n)$ values, the number of squares equivalent to a given square form a set of $\phi_1^2(n)$ squares. The total number of sets of non-equivalent squares is therefore $n^2\phi_1(n)\phi_2(n)$ or, if we consider symmetric squares only, $\phi_1(n)\ \phi_2(n)$. Thus, for example, the eight symmetric squares of three are in two non-equivalent sets:

$$\begin{array}{ccc} 2 & 7 & 6 \\ 9 & 5 & 1 \\ 4 & 3 & 8 \end{array} \qquad \begin{array}{ccc} 2 & 9 & 4 \\ 7 & 5 & 3 \\ 6 & 1 & 8 \end{array} \qquad \begin{array}{ccc} 8 & 1 & 6 \\ 3 & 5 & 7 \\ 4 & 9 & 2 \end{array} \qquad \begin{array}{ccc} 8 & 3 & 4 \\ 1 & 5 & 9 \\ 6 & 7 & 2 \end{array}$$

$$\begin{array}{ccc} 4 & 3 & 8 \\ 9 & 5 & 1 \\ 2 & 7 & 6 \end{array} \qquad \begin{array}{ccc} 6 & 1 & 8 \\ 7 & 5 & 3 \\ 2 & 9 & 4 \end{array} \qquad \begin{array}{ccc} 4 & 8 & 3 \\ 1 & 5 & 9 \\ 7 & 2 & 6 \end{array} \qquad \begin{array}{ccc} 6 & 7 & 2 \\ 1 & 5 & 9 \\ 8 & 3 & 4 \end{array}$$

From the fact that k and k' are prime to n it follows that if one of two equivalent squares is diabolic the other must be also. From the enumerative formula therefore the diabolic squares fall into $\phi_3(n)\phi_4(n)$ sets of non-equivalent squares. Thus the 32 diabolic squares of order 5 fall into 2 sets of 16 equivalent squares.

Any rearrangement of the rows or columns of a magic square will leave it magic since the sum in any row or column is thus left unaltered. A uniform step square, however, is not necessarily transformed in this way into a uniform step square. It appears easily that if the horizontal rows are transformed among themselves so that a row x goes into a row y where x and y are connected by the congruence relation $y \equiv \lambda x + \rho \pmod{n}$ where λ is prime to n, then this transformation H will throw a uniform step square into another uniform step square. It is found also that by this transformation the constants α and a of the old square are left unaltered while β and b are multiplied by λ. Similarly a transformation V of the same type on the columns throws a uniform step square into another uniform step square leaving the β and b unaltered and multiplying α and a by λ. It follows that

Those uniform step squares which are obtainable for a given square $\left(\begin{smallmatrix}\alpha & a\\ \beta & b\end{smallmatrix}\right)$ *by an H-transformation are of the form* $\left(\begin{smallmatrix}\alpha & a\\ k\beta & kb\end{smallmatrix}\right)$, *and those obtainable by a V-transformation are of the form* $\left(\begin{smallmatrix}k\alpha & ka\\ \beta & b\end{smallmatrix}\right)$.

The square $\left(\begin{smallmatrix}\alpha & a\\ \beta & b\end{smallmatrix}\right)$ and the square $\left(\begin{smallmatrix}k\alpha & ka\\ k'\beta & k'b\end{smallmatrix}\right)$ may be called equivalent under the transformations H and V. It appears that an H-transformation followed by a V-transformation can not produce a square equivalent under S and T transformations unless the parameter λ_1 of the H transformation equals the parameter λ_2 of the V transformation. The squares, therefore, that are equivalent under all four types of transformations are of the form $\left(\begin{smallmatrix}k\alpha & ka\\ k\beta & kb\end{smallmatrix}\right)$ and fall into sets of $\phi_1(n)$.

A third type of transformation is of use in the study of "nuclear" squares.* In this transformation the principal diagonals of the given square are made the row and column through the center of the new square. Any element of the old square lies in a cell which is the intersection of two diagonals which meet the principal diagonals in two elements x_1 and x_2. The corresponding cell in the new square for that element is found at the intersection of the row and column through those same elements x_1 and x_2. We proceed to set up the relations between the four numbers for the old and for the new cell.

Using the notions of elementary analytic geometry and calling the cell (1, 1) the origin, then the equation of the principal positive diagonal is found

* See John Willis, *Easy Methods of Constructing Magic Squares*, published by Percy Lund, Humphries and Co., London.

to be $y=-x+1$ and of the principal negative diagonal is $y=x$. Also the positive diagonal through the cell (A_1, B_1) has for its congruence $y-B_1 \equiv -(x-A_1)$. This meets the principal negative diagonal in the cell both of whose coördinates are $\frac{1}{2}(A_1+B_1)$. The negative diagonal through (A_1, B_1) meets the principal positive diagonal in the cell whose coördinates are $\frac{1}{2}(A_1-B_1+1)$, $\frac{1}{2}(B_1-A_1+1)$. Since now the cell (A_1', B_1'), the transform of (A_1, B_1), must project on the principal row and column in those cells in which the diagonals through (A_1, B_1) meet the principal diagonals in the first square, we have the coördinates (A_1', B_1') given by

$$A_1' = \frac{A_1+B_1}{2}, \quad B_1' = \frac{B_1-A_1+1}{2}.$$

Assuming, now, that the first square is symmetrical, we have

$$A_1 \equiv \frac{1}{2}(\alpha+a+1)+\alpha x+a\left[\frac{x-1}{n}\right],$$

$$B_1 \equiv \frac{1}{2}(\beta+b+1)+\beta x+b\left[\frac{x-1}{n}\right] \pmod{n};$$

therefore

$$A_1' \equiv \frac{1}{4}(\alpha+\beta+a+b+2)+\frac{\alpha+\beta}{2}x+\frac{a+b}{2}\left[\frac{x-1}{n}\right],$$

$$B_1' \equiv \frac{1}{4}(\beta+b-\alpha-a+2)+\frac{\beta-\alpha}{2}x+\frac{b-a}{2}\left[\frac{x-1}{n}\right] \pmod{n}.$$

The coefficients of the transformed square are therefore,

$$\begin{pmatrix} \frac{\alpha+\beta}{2}, & \frac{a+b}{2} \\ \frac{\beta-\alpha}{2}, & \frac{b-a}{2} \end{pmatrix}.$$

It is clear from these coefficients that this square is magic in the diagonals if the original is magic in the rows and columns, and magic in the rows and columns if the original is magic in the diagonals. Thus a diabolic square goes by this transformation into a diabolic square. Since the general effect is a sort of twist through an angle of forty-five degrees, we shall speak of this transformation as a rotation and denote it by R. We give as illustration the square of eleven $\left(\begin{smallmatrix}1&2\\2&5\end{smallmatrix}\right)$ and the transformed square $\left(\begin{smallmatrix}7&9\\6&7\end{smallmatrix}\right)$:

103	86	58	41	13	117	89	72	55	27	10
94	77	49	32	4	108	80	63	35	18	111
85	57	40	12	116	99	71	54	26	9	102
76	48	31	3	107	79	62	34	17	121	93
56	39	22	115	98	70	53	25	8	101	84
47	30	2	106	78	61	44	16	120	92	75
38	21	114	97	69	52	24	7	100	83	66
29	1	105	88	60	43	15	119	91	74	46
20	113	96	68	51	23	6	110	82	65	37
11	104	87	59	42	14	118	90	73	45	28
112	95	67	50	33	5	109	81	64	36	19

44	25	17	9	111	103	95	87	68	60	52
7	120	101	93	85	77	58	50	42	23	15
91	83	75	56	48	40	32	13	5	118	110
65	46	38	30	22	3	116	108	89	81	73
28	20	1	114	106	98	79	71	63	55	36
112	104	96	88	69	61	53	34	26	18	10
86	67	59	51	43	24	16	8	121	102	94
49	41	33	14	6	119	100	92	84	76	57
12	4	117	109	90	82	74	66	47	39	31
107	99	80	72	64	45	37	29	21	2	115
70	62	54	35	27	19	11	113	105	97	78

It should be noted that the principal positive diagonal might be taken as the row, or column, through the center of the new square, and that it might be taken to read either up or down for the column or right or left for the row. With these four ways of placing the principal positive diagonal will go also two ways of placing the principal negative diagonal in the new square, so that there should be eight different transformations of the type R for a given square. The transformations for these eight cases are given in the following table:

Positive diagonal reading up and down, negative diagonal right and left:

Down, right,

$$\begin{pmatrix} (\alpha+\beta)/2 & (a+b)/2 \\ (\beta-\alpha)/2 & (b-a)/2 \end{pmatrix},$$

Down, left,

$$\begin{pmatrix} -(\alpha+\beta)/2 & -(a+b)/2 \\ (\beta-\alpha)/2 & (b-a)/2 \end{pmatrix},$$

Up, right,

$$\begin{pmatrix} (\alpha+\beta)/2 & (a+b)/2 \\ (\alpha-\beta)/2 & (a-b)/2 \end{pmatrix},$$

Up, left,

$$\begin{pmatrix} -(\alpha+\beta)/2 & -(a+b)/2 \\ (\alpha-\beta)/2 & (a-b)/2 \end{pmatrix}.$$

Positive diagonal right and left, negative diagonal up and down:
Right, down,

$$\begin{pmatrix} (\alpha-\beta)/2 & (a-b)/2 \\ -(\alpha+\beta)/2 & -(a+b)/2 \end{pmatrix},$$

Left, down,

$$\begin{pmatrix} (\beta-\alpha)/2 & (b-a)/2 \\ -(\alpha+\beta)/2 & -(a+b)/2 \end{pmatrix},$$

Right, up,

$$\begin{pmatrix} (\alpha-\beta)/2 & (a-b)/2 \\ (\alpha+\beta)/2 & (a+b)/2 \end{pmatrix},$$

Left, up,

$$\begin{pmatrix} (\beta-\alpha)/2 & (b-a)/2 \\ (\alpha+\beta)/2 & (a+b)/2 \end{pmatrix}.$$

It will be observed that these squares are paired in four ways such that the first and third as well as the second and fourth elements are interchanged. They are also paired in four ways such that the first and second change their signs, or the third and fourth change their signs. It is important to see how the numbers in the squares themselves are related.

Certain special transformations on symmetric squares. It is of interest to note the effect of various special transformations on the constants α, β, a and b in symmetric squares. Thus it is easy to show that the two squares $\left(\begin{smallmatrix} \alpha & a \\ \beta & b \end{smallmatrix}\right)$ and $\left(\begin{smallmatrix} \beta & b \\ \alpha & a \end{smallmatrix}\right)$ are so related that the cell (A, B) becomes the cell (B, A). The effect, therefore, on the square is to make a reflection in the main diagonal. This result appears at once from the fundamental congruences defining the two squares. The same method will also serve to show that the two squares $\left(\begin{smallmatrix} \alpha & a \\ \beta & b \end{smallmatrix}\right)$ and $\left(\begin{smallmatrix} \alpha & -a \\ -\beta & -b \end{smallmatrix}\right)$ are so related that the second is the reflection of the first in its center. For by writing out the congruences for the cell (A, B) of the first square and for the corresponding cell (having the same x) in the second (A', B') it appears easily that $A+A' \equiv 1$ and $B+B' \equiv 1 \pmod{n}$. The effect is thus a rotation through two right angles about a line through the central cell perpendicular to the plane of the square. Simlarly $\left(\begin{smallmatrix} \alpha & a \\ -\beta & -b \end{smallmatrix}\right)$ is a reflection of $\left(\begin{smallmatrix} \alpha & a \\ \beta & b \end{smallmatrix}\right)$ in the row through the center.

The result of a transformation may sometimes be conveniently expressed by means of a relation between the numbers in corresponding cells of the two squares. For this purpose a number x is best represented in the form $x=kn+\rho$ where k is one of the numbers $0, 1, 2, 3, \cdots, n-1$ and ρ is one of the numbers $1, 2, 3, \cdots, n$. It will be observed that ρ is the smallest non-zero positive integer congruent to x modulo n, and that k is the number of the set (used in defining the transformation S) in which x appears. If the number x is divided by n, k is the quotient and ρ the remainder except for the case where x is a multiple of n, in which case k is the quotient less one and ρ is equal to n.

It is possible to show for the two symmetric squares $\left(\begin{smallmatrix}\alpha & a\\ \beta & b\end{smallmatrix}\right)$ and $\left(\begin{smallmatrix}-\alpha & a\\ -\beta & b\end{smallmatrix}\right)$ that $x=kn+\rho$ occupies the same cell in the first square that $y=kn+n+1-\rho$ does in the second square. For calling the cell in which x occurs (A_x, B_x) and the cell of the second square in which y occurs (A'_y, B'_y), we have

$$A_x = \frac{1}{2}(1+\alpha+a)+\alpha(kn+\rho-1)+a\left[\frac{kn+\rho-1}{n}\right],$$

or, from the definitions of k and ρ,

$$A_x = \frac{1}{2}(1+\alpha+a)+\alpha(\rho-1)+ak\,;$$

similarly

$$A'_y = \frac{1}{2}(1-\alpha+a)-\alpha(-\rho)+ak.$$

These two coördinates are manifestly the same. Similarly for the B's.

The two squares considered in the preceding paragraph are related by means of an S transformation on the numbers in the sets as follows:

$$s=\begin{pmatrix} 1, & 2, & 3, & \cdots, & n-2, & n-1, & n \\ n, & n-1, & n-2, & \cdots, & 3, & 2, & 1 \end{pmatrix}.$$

Any two numbers in corresponding cells in the two squares have the same k. There are n numbers which occupy the same cell in both squares.

It is possible to show for the two symmetric squares $\left(\begin{smallmatrix}\alpha & a\\ \beta & b\end{smallmatrix}\right)$ and $\left(\begin{smallmatrix}a & \alpha\\ b & \beta\end{smallmatrix}\right)$ that $x=kn+\rho$ occupies the same cell in the first that $y=(\rho-1)n+k+1$ does in the second. All that is necessary to prove this is to substitute the above value of x in the congruences belonging to the first square and the above value of y in the congruences belonging to the second.

These special theorems are of use in the study of "nuclear" squares to which we devote the next section.

Nodal and nuclear magic squares. It is possible to show that there are magic squares which are not uniform step squares and which are not derivable from such squares by any of the transformations which have just been discussed. Such a square, for example, is the following:

36	17	119	89	32	51	70	2	104	85	66
16	118	99	69	1	31	50	103	84	65	35
117	98	68	49	102	11	30	83	64	34	15
97	67	48	29	82	101	10	63	44	14	116
26	7	109	79	42	60	81	43	13	115	96
46	27	8	110	100	61	22	12	114	95	76
77	47	28	9	41	62	80	113	94	75	45
6	108	78	59	40	21	112	93	74	55	25
107	88	58	39	20	111	92	73	54	24	5
87	57	38	19	121	91	72	53	23	4	106
56	37	18	120	90	71	52	33	3	105	86

It will be observed that the middle column contains the numbers 51 and 62 which are both congruent to 7 modulo 11. It also contains the numbers 60 and 71 which are both congruent to 5. But in a uniform step square the numbers in any row or column form a complete series of incongruent numbers, and any one of the transformations discussed above does not disturb this property of uniform step squares. It follows that the above square is not a uniform step square nor derivable from such a square by any H, V, S or T transformation.

It will be further observed that the above square is *nuclear*; that is, it has a central square (of three rows and columns in this case) in which the rows and columns add up to the same amount. We extend the definition of *magic* to include squares such as

$$\begin{matrix} 42 & 60 & 81 \\ 100 & 61 & 22, \\ 41 & 62 & 80 \end{matrix}$$

which are not made up of numbers beginning with unity, but which do have the same sum in the rows and columns. We say then that the above square has a magic nucleus of three. It is possible to find squares with a magic nucleus of five, or seven or of m where m is odd.*

Certain types of uniform step squares can be manipulated so as to produce nuclear squares. Thus the above nuclear square is obtainable from the symmetrical uniform step square

* For examples of such squares see John Willis, loc. cit.

$$A \equiv 6 + x + \left[\frac{x-1}{11}\right], \qquad B \equiv 9 + x + 7\left[\frac{x-1}{11}\right] \pmod{11}.$$

The central square for this is found to be

$$\begin{matrix} 100 & 81 & 62 \\ 80 & 61 & 42 \\ 60 & 41 & 22 \end{matrix}.$$

We perform now an R transformation on this central square of 3 and get

$$\begin{matrix} 42 & 60 & 81 \\ 100 & 61 & 22 \\ 41 & 62 & 80 \end{matrix}.$$

The partial columns above the corners of this central square are then interchanged, and the partial rows to the left of the corners of this central square are also interchanged. The result is the nuclear square given above. The success of this method of getting nuclear squares is seen to depend on the fact that the central square of three is magic in its diagonals. In the same way we can obtain a nuclear square from any symmetrical uniform step square which has a central m-sided square which is magic in the diagonals. Since the square is symmetrical this common sum of all the diagonals must be $m(n^2+1)/2$. Such a square we will call a *nodal* square. To obtain a nuclear square from such a nodal square make an R transformation on the central square. Then the partial columns above and below the central square which are symmetrical in the nodal square must be so shifted that they lie in the same column. Thus, for example, the elements in the partial column above the upper left hand corner of the central square together with those in the partial column below the lower right hand corner of the central square add up to $(n-m)(n^2+1)/2$. If these two partial columns are then shifted to lie in the same vertical column the sum of all the elements in that column will be $(n-m)(n^2+1)/2+m(n^2+1)/2$ or $n(n^2+1)/2$, which is the right sum for a magic square. A similar shifting of the partial rows to the right and left of the central square will make the new square magic in all the rows. The problem of constructing nuclear squares by this method thus depends on the problem of finding symmetric uniform step squares which are nodal. It would seem possible to set up necessary and sufficient conditions on the constants of a uniform step square that it should be nodal, and perhaps from these conditions to obtain an enumeration of nodal squares. The author has not as yet been able to find such conditions, and the problem of enumeration of nodal squares seems to be of the same general type as the problem of finding the number of terms in a repeating decimal, or the number to which a given number belongs with respect to any modulus. He has suc-

ceeded, however, in proving certain theorems which have enabled him to compute the number of such squares for a given side without much difficulty. Thus he has been able to prove that if $\left(\begin{smallmatrix} a & a \\ \beta & b \end{smallmatrix}\right)$ is a nodal square we can write down at once the following list of nodal squares:

$$\left(\begin{smallmatrix} \alpha & a \\ \beta & b \end{smallmatrix}\right)\left(\begin{smallmatrix} a & \alpha \\ b & \beta \end{smallmatrix}\right)\left(\begin{smallmatrix} \beta & b \\ \alpha & a \end{smallmatrix}\right)\left(\begin{smallmatrix} b & \beta \\ a & \alpha \end{smallmatrix}\right)\left(\begin{smallmatrix} -\alpha & -a \\ \beta & b \end{smallmatrix}\right)\left(\begin{smallmatrix} -a & -\alpha \\ b & \beta \end{smallmatrix}\right)\left(\begin{smallmatrix} -\beta & -b \\ \alpha & a \end{smallmatrix}\right)\left(\begin{smallmatrix} -b & -\beta \\ a & \alpha \end{smallmatrix}\right)$$

$$\left(\begin{smallmatrix} -\alpha & a \\ -\beta & b \end{smallmatrix}\right)\left(\begin{smallmatrix} -a & \alpha \\ -b & \beta \end{smallmatrix}\right)\left(\begin{smallmatrix} -\beta & b \\ -\alpha & a \end{smallmatrix}\right)\left(\begin{smallmatrix} -b & \beta \\ -a & \alpha \end{smallmatrix}\right)\left(\begin{smallmatrix} \alpha & -a \\ -\beta & b \end{smallmatrix}\right)\left(\begin{smallmatrix} a & -\alpha \\ -b & \beta \end{smallmatrix}\right)\left(\begin{smallmatrix} \beta & -b \\ -\alpha & a \end{smallmatrix}\right)\left(\begin{smallmatrix} b & -\beta \\ -a & \alpha \end{smallmatrix}\right)$$

$$\left(\begin{smallmatrix} -\alpha & -a \\ -\beta & -b \end{smallmatrix}\right)\left(\begin{smallmatrix} -a & -\alpha \\ -b & -\beta \end{smallmatrix}\right)\left(\begin{smallmatrix} -\beta & -b \\ -\alpha & -a \end{smallmatrix}\right)\left(\begin{smallmatrix} -b & -\beta \\ -a & -\alpha \end{smallmatrix}\right)\left(\begin{smallmatrix} \alpha & a \\ -\beta & -b \end{smallmatrix}\right)\left(\begin{smallmatrix} a & \alpha \\ -b & -\beta \end{smallmatrix}\right)\left(\begin{smallmatrix} \beta & b \\ -\alpha & -a \end{smallmatrix}\right)\left(\begin{smallmatrix} b & \beta \\ -a & -\alpha \end{smallmatrix}\right)$$

$$\left(\begin{smallmatrix} \alpha & -a \\ \beta & -b \end{smallmatrix}\right)\left(\begin{smallmatrix} a & -\alpha \\ b & -\beta \end{smallmatrix}\right)\left(\begin{smallmatrix} \beta & -b \\ \alpha & -a \end{smallmatrix}\right)\left(\begin{smallmatrix} b & -\beta \\ a & -\alpha \end{smallmatrix}\right)\left(\begin{smallmatrix} -\alpha & a \\ \beta & -b \end{smallmatrix}\right)\left(\begin{smallmatrix} -a & \alpha \\ b & -\beta \end{smallmatrix}\right)\left(\begin{smallmatrix} -\beta & b \\ \alpha & -a \end{smallmatrix}\right)\left(\begin{smallmatrix} -b & \beta \\ a & -\alpha \end{smallmatrix}\right).$$

For certain values of n, of course, some of these may not be distinct. He has also been able to show that in a nodal square with a node of three, the entries in the columns and rows of the node are in actual arithmetical progression. This seems, experimentally, to hold for nodes of five and so on, but a proof is lacking for this. This theorem makes the enumeration of squares with a node of three a fairly simple matter even though an enumeration formula does not present itself. The number of nodal squares with a node of three for the first few values of n are as follows:

n	number of squares
5	8
7	120
9	128
11	1648
13	3272
15	512
17	11712
19	19520
21	3072
23	46072
25	36424
27	12800

Since there are eight ways of performing the transformation R on the central node and four ways of adjusting the partial columns above and below and to the right and left of the node to turn the square into a nuclear square, the above numbers must be multiplied by 32 for the number of nuclear squares with a nucleus of three obtainable by the above method. These nuclear squares are also still transformable into magic squares by the

transformations H and V on the rows and columns, in number $(n!)^2$. The transformations S and T are however not available for squares of this sort.

The study of nodal squares brings to light many curious and interesting theorems of the following type:

The square $\left(\begin{smallmatrix}1&1\\1&2\end{smallmatrix}\right)$ is nodal for all values of n. The square $\left(\begin{smallmatrix}1&1\\1&7\end{smallmatrix}\right)$ is nodal for all values of n prime to 6. The square $\left(\begin{smallmatrix}1&1\\1&5\end{smallmatrix}\right)$ is nodal for $n=4k+1$, but not for $n=4k-1$. The square $\left(\begin{smallmatrix}1&1\\1&8\end{smallmatrix}\right)$ is nodal for numbers of the form $14k-1$ and for no others. The square $\left(\begin{smallmatrix}4&3\\3&5\end{smallmatrix}\right)$ is never nodal for any value of n, etc.

UNIVERSITY OF CALIFORNIA,
BERKELEY, CALIF.

CERTAIN PROBLEMS RELATED TO THE CUTTING OF A SIMPLY CONNECTED PLANE REGION BY A CONTINUUM*

BY

WALLACE ALVIN WILSON

1. If R is a simply connected plane region whose frontier F is bounded and C is a simple arc lying in R save for its end points, it is well known that C enjoys the following properties:

(a) $R-R\cdot C$ consists of two simply connected regions R_1 and R_2;

(b) C is a part of the frontiers of both R_1 and R_2;

(c) $F+C$ is the union of the frontiers of R_1 and R_2.

Nothing could be simpler than this theorem, but when one desires to use it, certain questions arise. For instance, what precisely are the frontiers of R_1 and R_2? If F is the union of two continua H and K, can C be so drawn that its ends lie on two components of $H\cdot K$? If so, will the frontiers of the two regions be $C+H$ and $C+K$? These questions are readily answered in the case that F is a simple closed curve, but the answers in this case do not always remain correct when F is more complicated. In particular, a negative answer to the second question suggests the substitution of a more general continuum for C and this in turn leads to a re-examination of the truth of the resulting extension of the theorem.

Such an extension has been made elsewhere† and it is intended in the present article to examine further these questions and certain related topics. The first and third of the questions suggested above are discussed in §§3–6. The remainder of the paper deals with the situation arising when the arc C is replaced by a special kind of irreducible continuum and with related questions of accessibility. In particular, it is shown that under certain broad conditions the second and third of the above questions can be answered in the affirmative if the arc C is replaced by a slightly more general continuum.

2. **Notation.** In addition to the common notation of the aggregate theory, the following will be used.

The whole plane is denoted throughout by Z. If F is a continuum, a component of $Z-F$ whose frontier is F is called a *principal* component; a component of $Z-F$ whose frontier is a proper part of F is called *secondary*. A

* Presented to the Society, October 27, 1928; received by the editors August 20, 1928.

† W. A. Wilson, *On irreducible cross-cuts of plane simply connected regions*, American Journal of Mathematics, vol. 51, pp. 19–30.

frontier set whose complement has at least two principal components is called *regular*.

The statement that C is a continuum irreducible between the sets A and B does not imply that C contains the whole of either A or B.

The symbol $V_\delta(a)$ denotes the set of points whose distance from the point a is less than δ.

If m and n are two points (or point sets) in the plane separated by the closed set C, we say that C is an $S(m, n)$. If C is an $S(m, n)$, but no proper closed subset of C is an $S(m, n)$, we say that C is an *irreducible* $S(m, n)$.

3. We shall begin with the question of the nature of the frontiers of the regions R_1 and R_2 referred to in §1. It is clear that this depends upon the degree of complexity of F and that a satisfactory solution is scarcely possible unless restrictions are imposed on F. In a discussion of this kind the principal property of a simple closed curve is that it is the common frontier of its complementary domains; hence the natural generalization of such a curve is the regular frontier. The principal properties of such frontiers which will be used have been given by C. Kuratowski* and the author.† The following theorem is also convenient.

THEOREM. *Let F be the frontier of a simply connected region R and let S be any other component of $Z-F$. If the frontier of S is bounded, it is regular.*

Let f be the frontier of S and let m and n be points of R and S, respectively. As $f \subset F$, f is an irreducible $S(m, n)$. Then f is a regular frontier (see Kuratowski, loc. cit., p. 133).

4. THEOREM. *Let R be a simply connected region whose frontier F is bounded and regular. Let C be a simple arc joining the points a and b of F and contained in $R+a+b$. Let $R \cdot C$ divide R into R_1 and R_2 with frontiers F_1 and F_2, respectively. Then F contains either* 0, 1, *or* 2 *proper sub-continua irreducible between a and b. In the first case $F_1=F_2=C+F$; in the second case $F_2=C+F$, $F_1=C+f_1$, where $f_1 \subset F$ and is irreducible between a and b; in the third case $F_1=C+f_1$ and $F_2=C+f_2$, where f_1 and f_2 are proper sub-continua of F irreducible between a and b.*

If F is irreducible between a and b, $C+F$ has two principal complementary domains by Rosenthal's theorem.‡ These obviously coincide with R_1 and R_2.

---

* C. Kuratowski, *Sur les coupures irréductibles du plan*, Fundamenta Mathematicae, vol. 6, pp. 130–145.

† W. A. Wilson, *On bounded regular frontiers in the plane*, Bulletin of the American Mathematical Society, vol. 34, pp. 81–90.

‡ A. Rosenthal, *Teilung der Ebene durch irreduzibele Kontinua*, Sitzungsberichte der Münchener Akademie, 1919, p. 106.

If f_1 is a proper sub-continuum of F irreducible between a and b, let S_1 and S_2 be the principal complementary domains of $C+f_1$. Since F is regular, $M=F-f_1$ is connected and lies in one of these, say in S_2. Then $F\cdot S_1=0$ and $S_1 \subset R$. It readily follows that $S_1=R_1$ or $S_1=R_2$, say the former. Since $F_1+F_2=C+F$, this shows that $F_2 \supset M$. Let $N=F\cdot F_2$. Let m be a point of R_2 and n a point of some principal component of $Z-F$ other than R. Since F is regular, N is not an $S(m, n)$ unless $N=F$. Then $F_2=C+F$ and a repetition of the first part of this paragraph shows that another proper sub-continuum of F irreducible between a and b is impossible, when $N=F$.

If $N \neq F$, N does not contain f_1 since $N \supset M$. As neither N nor C is an $S(m, n)$, $N\cdot C$ is disconnected by Janiszewski's theorem* and N contains a and b. For the same reason N contains a sub-continuum f_2 irreducible between a and b. As $f_1 \neq f_2$, the reasoning of the previous paragraph shows that $F_2=C+f_2$.

Finally, the existence of a third proper sub-continuum of F irreducible between a and b is impossible by virtue of the second paragraph of the proof, as C divides R into but two regions.

5. A consequence of the theorem of §3 is that §4 applies to any simply connected plane region whose frontier is bounded and a part of the frontier of some other region.

It is natural to expect that, if the frontier F of R is regular, then the frontiers of the regions R_1 and R_2 into which C divides R are regular. The second case of the above theorem shows that this is not true. For, if m, n, and p are points of R_2, R_1, and $Z-\overline{R}$, respectively, the frontier of R_2, which is $F+C$, is neither an irreducible $S(m, n)$ nor an irreducible $S(m, p)$.

A further consequence of the above theorem is that, if a and b are two points of a bounded regular frontier F accessible by a simple arc from one of the components of $Z-F$, there are at most two sub-continua of F irreducible between a and b. This need not be true if either a or b is not arcwise accessible.

To determine precisely which of the three cases mentioned in the theorem applies, we need to know the nature of F and the location of a and b on F. The various possibilities may be classified briefly as follows. The notation is that of §4.

(I) *F is indecomposable.*

---

* Z. Janiszewski, *Sur les coupures du plan faites par des continus*, Prace Matematyczno-Fizyczne, vol. 26, Theorem A. If P and Q are bounded closed sets, if $P\cdot Q$ is connected, and neither P nor Q is an $S(m, n)$, then $P+Q$ is not an $S(m, n)$. See also S. Straszewicz, Fundamenta Mathematicae, vol. 4, p. 129.

(a) If a and b are not on the same constituent of F, $F_1=F_2=C+F$.

(b) If a and b are on the same constituent of F, $F_1=C+f_1$ and $F_2=C+F$.

Examples of both these cases are easily constructed by using the methods of K. Yoneyama†.

(II) *F is the union of two indecomposable continua.* In this case F is the union of two continua H and K, $H\cdot K=\sum_1^n\alpha_i$, where $n\geqq 2$ and depends upon the number of principal components of $Z-F$, each α_i is a closed set, and both H and K are irreducible between each pair of sets $\{\alpha_i\}$.‡ The various possibilities are as follows.

(a) If a and b are on different sets α_i, $F_1=C+H$ and $F_2=C+K$. (For an example see Ex. 2 of Wilson, loc. cit. II, taking $C=cd$.)

(b) If a and b are on the same component of some α_i, $F_1=C+F$ and $F_2=C+f$, where $f\subset\alpha_i$.

(c) If a and b are on the same set α_i, but on different components, both H and K are irreducible between them, $F_1=C+H$, and $F_2=C+K$.

For we then have two sub-continua f_1 and f_2 of F irreducible between a and b, $F_1=C+f_1$, and $F_2=C+f_2$, which gives the contradiction $F_1+F_2\neq C+F$ unless $f_1=H$ and $f_2=K$. As a corollary it follows that, if two components of any α_i are arcwise accessible from R, both H and K are irreducible between them.

(d) If a is on some α_i and b on no α_i, there are three possibilities. If H is reducible between a and b, $F_1=C+f_1$, where $f_1\subset H$, and $F_2=C+F$. If H is irreducible between a and b, but reducible between b and some α_j, $F_1=C+H$ and $F_2=C+K+f_1$, where $f_1\subset H$. If H is irreducible between b and every α_i, $F_1=C+H$ and $F_2=C+F$.

(e) If neither a nor b is on any α_i, we may have any of the three cases in §4.

(III) *F is neither indecomposable nor the union of two indecomposable continua.* In this event F is not irreducible between a and b and we may have either Case 2 or Case 3 of §4.

6. THEOREM. *Let H and K be bounded continua, $H\cdot K\neq 0$ and not connected, $F=H+K$, and $H^*=F-K$ and $K^*=F-H$ be connected. Let R be a component of $Z-F$, let a and b be points on different components of $H\cdot K$, and let C be a simple arc joining a and b and contained in $R+a+b$. Let F_1 and F_2 be the frontiers of the components of $R-R\cdot C$. Then $F_1=C+h$ and $F_2=C+k$, where h and k are sub-continua of H and K, respectively, joining a and b.*

† Tôhoku Mathematical Journal, vol. 12, pp. 60–62.

‡ See W.A. Wilson, loc. cit. II, p. 86.

Let R_1 and R_2 be the components of $R-R\cdot C$ and let S_1 and S_2 be the components of $Z-(C+H)$ whose frontiers contain C. Let S_2 be the one which contains the connected set K^*. Since $H\cdot S_1=0$, $(H+K^*)\cdot S_1=F\cdot S_1=0$ and $S_1\subset R$. If $S_1\cdot R_1\neq 0$, $S_1\cdot R_2=0$, since $C\cdot S_1=0$; and vice versa. Let then $S_1\subset R_1$. But then $S_1=R_1$, since $R_1\cdot(C+H)=0$.

Now let m be a point of R_1 and n one of R_2. Let F_1 be the frontier of R_1 and $h=H\cdot F_1$; then $F_1=C+h$. Either h contains a continuum joining a and b, or $h=h_1+h_2$, where $h_1\cdot h_2=0$, $a\subset h_1$, $b\subset h_2$, and h_1 and h_2 are closed. As neither h_1, h_2, nor C is an $S(m, n)$, but F_1 is, the latter alternative is impossible by Janiszewski's theorem, loc. cit. As a and b are on different components of $H\cdot K$, $h\cdot H^*\neq 0$.

Now let h' be a sub-continuum of h irreducible between a and b. If h is not a continuum, $h=h_1+h_2$, where h_1 and h_2 are closed and $h_1\cdot h_2=0$. As $h\supset h'$, either $h'\subset h_1$ or $h'\subset h_2$, say the former. But then $F_1=C+h=(C+h_1)+h_2$ is not a continuum, for $h_2\cdot(C+h')=0$. This is a contradiction.

In like manner there is a component T of $Z-(C+K)$ which is identical with R_1 or R_2, and whose frontier contains C and points of K^*. As $F_1\subset C+H$, $T\neq R_1$. Hence $T=R_2$ and $F_2=C+k$, where k is a sub-continuum of K.

COROLLARY. *If in addition to the above hypotheses F is the frontier of R and the sum $H+K$ is irreducible*†, *$F_1=C+H$ and $F_2=C+K$.*

For $F_1+F_2=F+C$ and hence $F_1\supset H^*$ and $F_2\supset K^*$. But $\overline{H}^*=H$, since $H+K$ is an irreducible sum. Then $F_1=C+H$ and $F_2=C+K$.

Remarks. A consequence of the above theorem is that not more than two components of $H\cdot K$ contain points arcwise accessible from R. (See also §12.)

Very simple examples show that the omission of the hypothesis that H^* and K^* are connected renders the theorem invalid and that the hypothesis that $H+K$ is an irreducible sum cannot be omitted from the corollary.

If no statements are made regarding a partition of F into sub-continua and a and b are any two points on F, we can only say that the frontier of each of the regions R_1 and R_2 contains C and a sub-continuum of F irreducible between a and b. This follows readily from Janiszewski's theorem, loc. cit. The two irreducible continua involved may be identical, the simplest example being when F is a straight line segment. In the special case that a and b lie on the frontier of some component of $Z-F$ other than R, we can by virtue of the theorem of §3 make more precise statements regarding these irreducible continua in accordance with the classification of §5.

† This means that there is no proper sub-continuum H' of H (or K' of K) such that $F=H'+K$ (or $F=H+K'$).

7. The theorem of §6 is not applicable unless two components of $H \cdot K$ contain points arcwise accessible from R, and examples not having this property are easily constructed. If in such a case we desire to separate R by a continuum C into two domains whose frontiers are $C+H$ and $C+K$, continua more general than the simple arc must be used.

We first note that, if m and n are points of the two domains just referred to, C separates m from n in R. If, furthermore, C is irreducible with respect to this property and is not itself an $S(m, n)$, it has been shown elsewhere† that C is an irreducible continuum, which may be indecomposable. It seems desirable, however, in the interests of simplicity, to confine the discussion to the case where C is neither indecomposable nor the union of two indecomposable continua. A further justification of this slight loss in generality is the easily established fact that, if C has the separation properties above mentioned and meets the frontier of R in two closed sets A and B such that $A \cdot B = 0$, there is a continuum C' joining A to B and separating m from n in R irreducibly, which is decomposable into more than two continua.

To save words a decomposable continuum which is not the union of two indecomposable continua will be called *semi-separable*. It is easily shown that, if C is semi-separable and irreducible between the closed subsets α and β, then $C = A+E+B$, where A, B, and E are continua, $A \cdot B = 0$, $\alpha \subset A$, $\beta \subset B$, and E is irreducible between A and B.

8. THEOREM. *Let C be a bounded semi-separable continuum irreducible between the subsets α and β. Let no sub-continuum of C separate a point of α from one of β. Then $Z-C$ has exactly one principal component.*

By §7, $C = A+E+B$, where A, B, and E are continua irreducible between α and E, β and E, and A and B, respectively. Let a and b be points of α and β, respectively, and let $\epsilon > 0$ be so small that $(E+B) \cdot V_\epsilon(a) = 0$ and $(A+E) \cdot V_\epsilon(b) = 0$. Set $\delta_i = \epsilon/2^i$ and let V_i and W_i denote the sets of points whose distances from a and b, respectively, are less than δ_i.

Set $C_i = C(Z - V_i - W_i)$. Then C_i is bounded and closed and is not an $S(a, b)$, since by hypothesis none of its components is an $S(a, b)$. Hence there is a simple arc from a to b not cutting C_i. Let L_i be a segment of this joining $\overline{V}_i$ to $\overline{W}_i$. Then $C \cdot L_i = 0$ and L_i lies in a component R of $Z-C$.

Now suppose that for some $j \neq i$, the arc L_j were in a different component S of $Z-C$. Let x and y be points of L_i and L_j, respectively. Then C is an $S(x, y)$. But it is obvious that neither $A+E$ nor $E+B$ is an $S(x, y)$ on account of the choice of ϵ, and $(A+E) \cdot (E+B) = E$, a continuum. This is a

† See Wilson, loc. cit. I, §11.

contradiction by the footnote on page 554. Hence $S=R$; i.e., for every i, L_i lies in the same component R of $Z-C$. Then R contains points in every V_i and every W_i. As $\delta_i \to 0$, the frontier of R must contain a and b. Since C is irreducible between a and b, it follows that C is the frontier of R.

Obviously this can be true of no other component of $Z-C$.

COROLLARY. *Let C satisfy the hypotheses of the above theorem and $C=A+E+B$, where A, E, and B are continua and E is irreducible between A and B. If S is a secondary component of $Z-C$, its frontier is a part of $A+E$ or of $E+B$.*

Let R be the principal component of $Z-C$, let m be a point of R, and n one of S. If F is the frontier of S, F is an $S(m, n)$. The corollary is obviously true unless F contains points in both A and B. Then $E+F$ is an $S(m, n)$. But $(E+F\cdot A)(E+F\cdot B)=E$. Hence either $E+F\cdot A$ or $E+F\cdot B$ is an $S(m, n)$, say the former. Let y be a point of $F\cdot B$ not in $E+A$. Then $(E+A)\cdot V_\delta(y)=0$ for some $\delta>0$. But $R\cdot V_\delta(y)\neq 0$ and $S\cdot V_\delta(y)\neq 0$. Hence there is a continuum joining m and n and not meeting $E+F\cdot A$. Thus we have a contradiction unless $F\cdot B=0$.

Remarks. The condition that no sub-continuum E of C may separate a point of α from one of β is a necessary one. For in the contrary case E separates every point of α from β, as C is irreducible between α and β, and then there would be no principal component of $Z-C$.

The theorem does not remain valid if the word "semi-separable" is omitted. For in Ex. 1, p. 87 of Wilson, loc. cit. II, is given an irreducible continuum whose complement has no principal component. This continuum is the union of two indecomposable continua, but in other respects satisfies the hypotheses of the theorem.

9. THEOREM. *Let $F=C+K$, where C and K are bounded continua. Let $C\cdot K$ be the sum of two closed sets without common points and let C be semi-separable and irreducible between them. Then C is a proper part of the frontiers of exactly two components of $Z-F$.*

By §13 of Wilson, loc. cit. I, there are at least two components of $Z-F$ such that C is a proper part of their frontiers. Now $C-C\cdot K$ is connected and lies in one component R of $Z-K$. Then there are at least two points m and n of R such that C is an irreducible $S(m, n)$ in R. On the other hand C is not an $S(m, n)$ by §8 and hence §11 of Wilson, loc. cit. I, is applicable. This gives the theorem, since C is semi-separable and hence not irreducible between each pair of three or more closed sets.

COROLLARY 1. *Let R be a simply connected region whose frontier F is bounded. Let α and β be closed subsets of F without common points, let C be a bounded continuum contained in $R+\alpha+\beta$, and let C be semi-separable and irreducible between α and β. Then C is a proper part of the frontiers of exactly two of the simply connected regions into which it divides R.*

COROLLARY 2. *Let R be a simply connected region whose frontier F is bounded. Let α and β be sub-continua of F without common points, let C be a bounded continuum contained in $R+\alpha+\beta$ and containing $\alpha+\beta$, and let C be semi-separable and irreducible between α and β. Let C separate no two points of $F-C\cdot F$. Then there are exactly two components R_1 and R_2 of $R-R\cdot C$ whose respective frontiers F_1 and F_2 contain C and $F_1+F_2=C+F$.*

The corollary is obvious unless there is another component $R_i(i>2)$ of $R-R\cdot C$. If so, let F_i be its frontier and $H_i=C\cdot F_i$. Then $H_i\neq C$ and $H_i=K_i+L_i$, where K_i and L_i are closed and $K_i\cdot L_i=\alpha\cdot L_i=\beta\cdot K_i=0$. Let x be a point of R_1 and y one of R_i. Then F_i, and hence $H_i+\alpha+\beta+F$, is an $S(x, y)$. Since $(K_i+\alpha+F)\,(L_i+\beta+F)=F$, either $K_i+\alpha+F$ or $L_i+\beta+F$ is an $S(x, y)$, say the former. Now F is not an $S(x, y)$, but $(K_i+\alpha)\,(\alpha+F)=\alpha$, a continuum. Thus $K_i+\alpha$ is an $S(x, y)$.

This shows that R_i is a part of a secondary component of $Z-C$, whereas R_1 is a part of the principal component. If F_i contained points of $F-C\cdot F$, it would follow that C separates $F-C\cdot F$, contrary to the hypothesis. Hence $F-C\cdot F\subset F_1+F_2$ and consequently $C+F=F_1+F_2$.

COROLLARY 3. *Let H and K be bounded continua, $H\cdot K\neq 0$, $F=H+K$, and $H^*=F-K$ and $K^*=F-H$ be connected. Let R be a component of $Z-F$, let α and β be closed subsets of two different components of $H\cdot K$, and let C be a bounded semi-separable continuum contained in $R+\alpha+\beta$ and irreducible between α and β. Let R_1 and R_2 be the components of $R-R\cdot C$ whose frontiers F_1 and F_2 contain C. Then $F_1\subset C+H$, $F_2\subset C+K$, $F_1\cdot H^*\neq 0$, and $F_2\cdot K^*\neq 0$.*

This follows by the reasoning of the first two and last paragraphs of §6.

COROLLARY 4. *Let R be a simply connected region whose frontier F is the irreducible sum of two bounded continua H and K, such that $H^*=F-K$ and $K^*=F-H$ are connected and $H\cdot K$ is not connected. Let α and β be closed subsets of two different components of $H\cdot K$ and let C be a bounded continuum contained in $R+\alpha+\beta$ which is semi-separable and irreducible between α and β. Then there are two components of $R-R\cdot C$ whose respective frontiers are $C+H$ and $C+K$.*

By Corollary 3 there are two components R_1 and R_2 whose frontiers F_1 and F_2 are parts of $C+H$ and $C+K$, respectively, and contain C. If there is

any other component R_i of $R-R\cdot C$, it can be shown as in Corollary 2 that any point of R_i is separated from one of R_1 by the union of a proper sub-continuum of C and one of the components of $H\cdot K$. As H^* is connected and $F_1\cdot H^*\neq 0$, the frontier F_i of R_i contains no point of H^*. Then, if U is the union of the sets F_i for $i>2$, $\overline{U}$ contains no point of H^*. Hence $H^*+K^*\subset F_1+F_2$. Since $F=H+K$ is an irreducible sum, $\overline{H^*}=H$ and $\overline{K^*}=K$. Combining these statements gives the desired result.

Remarks. The requirement in Corollary 2 that C must not separate $F-C\cdot F$ is not trivial or superfluous. To see this, let F be the union of a circumference M and a segment ab of a radius, where a lies on M and b within M. Let R be the bounded component of $Z-F$, $\alpha=a$, and β be another point on M. Let C be the union of a circumference N lying in R and tangent to N at a and a ray in R starting at β and approaching N asymptotically. Then $F_1+F_2=C+M$, but not $C+F$.

In regard to Corollary 4 it will be shown later by a theorem on accessibility that α and β cannot lie on more than two components of $H\cdot K$ and be subsets of a continuum C in $R+\alpha+\beta$. This corollary is false unless both H^* and K^* are connected.

10. The notion of arc-wise accessibility has been generalized† in a form equivalent to the following definition.

If F is a bounded continuum and R is a component of $Z-F$, the sub-continuum f of F is called accessible from R if for each point a of R there is a bounded continuum C such that $C\subset R+f$; $C\supset a$; $C\cdot f\neq 0$; $C\cdot F\subset f$.

The choice of a is obviously immaterial. It may well happen that no *point* of f is accessible from R, either by a line or a continuum. As an example, let F consist of a circumference f, two spirals approaching f asymptotically, and a chord joining the spirals, and let R be one of the principal components of $Z-F$. Then f, but no point of f, is accessible from R. The following is a special case of a theorem announced in the abstract referred to above.†

THEOREM. *Let F be a bounded continuum and R a component of $Z-F$. Let f be a sub-continuum of F accessible from R and a be any point of R. Then there is a ray contained in R whose initial point is a and whose improper limiting points lie on f.*

11. THEOREM. *Let H and K be bounded continua, $H\cdot K\neq 0$, $F=H+K$, and $H^*=F-K$ and $K^*=F-H$ be connected. If R is a component of $Z-F$, there is no continuum C contained in $R+H\cdot K$ which has points on more than two components of $H\cdot K$.*

† G. T. Whyburn, *A generalized notion of accessibility*, Bulletin of the American Mathematical Society, vol. 34, p. 429 (abstract).

Let us assume that R is bounded; a similar proof holds if R is unbounded. Now suppose the existence of a continuum C contained in $R+H\cdot K$ and having points on three components α_1, α_2, and α_3 of $H\cdot K$. Then $H\cdot K=\sum_1^3 A_i$, where each A_i is closed, no two have common points, and $\alpha_i \subset A_i$.

Let P be a polygon in R whose interior I contains a point a of C and which is so large that no component of $C-C\cdot I$ connects any pair of the sets $\{A_i\}$. For each i let C_i be a sub-continuum of C irreducible between P and A_i and let L_i be a simple arc within P joining a to C_i; let $L_i\cdot L_j=a$ if $i\neq j$. Set $D_i=L_i+C_i$.

Now D_1+D_2 is semi-separable and irreducible between A_1 and A_2. By §9, Corollary 1, D_1+D_2 divides R into several domains of which there are two, R_1 and R_2, whose frontiers F_1 and F_2 contain D_1+D_2. By Corollary 3 of §9, $F_1\subset H+D_1+D_2$ and $F_2\subset K+D_1+D_2$.

As D_3 is irreducible between a and A_3 and $D_3\cdot(D_1+D_2)=a$, D_3 lies in $\overline{R}_1$ or $\overline{R}_2$, say in $\overline{R}_2$. Applying the same references again, we see that there are two components R_{21} and R_{22} of $R_2-R_2\cdot D_3$ whose frontiers contain D_1+D_3 and D_2+D_3, respectively, and a part of K^*. (For, as shown in the demonstration of §9, Corollary 3, the frontier of R_{21} contains a sub-continuum of D_1+K joining a and A_3.) On the other hand there are two components S_1 and S_2 of $R-R\cdot(D_1+D_3)$ whose frontiers contain D_1+D_3 and parts of K and H, respectively. One of these, say S_1, equals R_{21}. Then $S_2\supset R_1$, since R_1 borders the arc L_1. Also $S_2\supset R_{22}$, since R_{22} borders the arc L_2. This is a contradiction, since S_2 does not have any frontier points on K^*, but R_{22} does. Hence the theorem is proved.

12. THEOREM. *Let F be the union of two bounded continua H and K such that $H\cdot K$ is not connected and $H^*=F-K$ and $K^*=F-H$ are connected. Let R be a component of $Z-F$ having frontier points on both H^* and K^*. Then exactly two components of $H\cdot K$ are accessible from R.*

By a theorem of Knaster and Kuratowski† there is a bounded continuum P which is an irreducible $S(H^*, K^*)$ and of course contains points of each component of $H\cdot K$.

Let $M=P\cdot R$. If $\overline{M}$ had points on three components α, β, γ of $H\cdot K$, we would have a contradiction. For let A, B, C be components of M having limiting points on α, β, γ, respectively. Join A to B and C by arcs D and E within R. Then $\overline{A}+\overline{B}+\overline{C}+D+E$ would be a continuum in $R+H\cdot K$ having points on more than two components of $H\cdot K$, contrary to §11.

Suppose now that $\overline{M}$ had points on only one component α of $H\cdot K$.

† B. Knaster and C. Kuratowski, *Sur les ensembles connexes*, Fundamenta Mathematicae, vol. 2, p. 236.

Let $N=P-M$. Then $Q=\overline{M}+\alpha+N$ is an $S(H^*, K^*)$. The set $N+\alpha$ is not an $S(H^*, K^*)$, since R has frontier points on both H^* and K^*. Since there is at least one other component of $Z-F$ having this property, $\overline{M}+\alpha$ is not an $S(H^*, K^*)$. As $(\overline{M}+\alpha)\cdot(\alpha+N)=\alpha$, a continuum, we have a contradiction.

Thus $\overline{M}$ has points on two components α and β of $H\cdot K$. If $\overline{M}$ does not contain a continuum irreducible between α and β, $\overline{M}=M_1+M_2$, where M_1 and M_2 are closed and $M_1\cdot M_2=M_1\cdot\beta=M_2\cdot\alpha=0$. Then $N+\alpha+\beta$ is a continuum, since P is a continuum. As M_1 and M_2 contain points in R, neither $M_1+N+\alpha+\beta$ nor $M_2+N+\alpha+\beta$ is an $S(H^*, K^*)$ by the argument of the previous paragraph. But $(M_1+N+\alpha+\beta)\cdot(M_2+N+\alpha+\beta)=N+\alpha+\beta$, a contradiction. Hence $\overline{M}$ contains a continuum irreducible between α and β and meeting $F-(\alpha+\beta)$ nowhere. This proves the theorem.

Remark. If in the above the components of $H\cdot K$ accessible from R are points, they are arcwise accessible from R. This follows at once from §10.

13. Various conclusions can be deduced by combining the theorems of §§9–12. The most interesting is perhaps the following.

Suppose that R is a simply connected region whose frontier F is a bounded continuum which is the union of two continua H and K such that $H^*=F-K$ and $K^*=F-H$ are connected. If $H\cdot K$ is not connected, it follows from §§10 and 12 that there are exactly two components α and β of $H\cdot K$, each of which is accessible from R by a ray. Then there is a bounded connected set C^* which is contained in R and has these properties: C^* is the homeomorphic image of a simple arc and its limiting points form two continua of condensation lying on α and β, respectively. As $C=\bar{C}^*$ is a continuum irreducible between α and β, we may apply §9, Corollary 3. This, together with the fact that $C-C\cdot F$ does not cut the plane tells us that C divides R into two domains R_1 and R_2, whose frontiers F_1 and F_2 contain C and parts of H and K, respectively. If, furthermore, $F=H+K$ is an irreducible sum, we know by §9, Corollary 4, that $F_1=C+H$ and $F_2=C+K$. Thus we have a very simple generalization of the division of a region bounded by a simple closed curve into two regions by a simple arc, applicable to very general frontiers and involving only the replacement of the simple arc by a continuum essentially equivalent to it.†

† Since the submission of this paper to the editors, there has been published in the Fundamenta Mathematicae, vol. 12, pp. 214–239, an article by C. Kuratowski (*Sur la séparation d'ensembles situés sur le plan*) containing a series of theorems, one of which (p. 227, Cor. 1) can be used to obtain this result.

YALE UNIVERSITY,
NEW HAVEN, CONN.

AN APPLICATION OF THE CALCULUS OF VARIATIONS TO BOUNDARY VALUE PROBLEMS*

BY

A. O. HICKSON

INTRODUCTION

In a paper published in 1910 R. G. D. Richardson considered the Jacobi condition of the calculus of variations and, by means of well known theorems from integral equation theory and the calculus of variations, proved certain existence theorems and oscillation properties for the solutions of a restricted type of ordinary linear differential equations of the second order which contain a parameter.†

Having set up the integral

$$D(u) = \int_0^1 [p(x)u'^2 - q(x)u^2]dx \qquad \left(u' = \frac{du}{dx}\right),$$

where p and q are analytic functions of x in the interval $0 \leqq x \leqq 1$, p being positive and q nowhere positive within the interval, he considered the problem of finding in a class of arcs satisfying the conditions

$$(1) \qquad u(0) = 0, \quad u(1) = 0, \quad \int_0^1 k(x)u^2dx = 1,$$

where k is an analytic function of x on the interval $0 < x < 1$, one for which $D(u)$ takes on a minimum value. The Euler-Lagrange differential equation of this isoperimetric problem was found to be

$$(2) \qquad L_1(u) = \frac{d}{dx}(pu') + qu + \lambda ku = 0.$$

The function k was first supposed to be not negative‡ for $0 \leqq x \leqq 1$, and it then followed from the Hilbert theory of the integral equation corresponding to the boundary value problem defined by (2) and the first two equa-

* Presented to the Society, April 6, 1928; received by the editors November 5, 1928.

† R. G. D. Richardson, *Das Jacobische Kriterium der Variationsrechnung und die Oszillationseigenschaften linearer Differentialgleichungen 2. Ordnung*, Mathematische Annalen, vol. 68 (1910), p. 279.

‡ The case where k changes sign in the interval was also treated.

tions of (1), that there exists a definitely increasing sequence of values $\lambda_i (i = 1, 2, \cdots)$ for each of which the equation $L_1(u) = 0$ possesses a unique solution U_i which satisfies the conditions imposed in the isoperimetric problem. Again making use of the theory of integral equations he was able to prove that in the class of functions $u(x)$ satisfying the conditions

$$u(0) = 0, \qquad u(1) = 0, \qquad \int_0^1 k(x)u^2 dx = 1,$$

$$\int_0^1 kU_1 u dx = 0, \qquad \int_0^1 kU_2 u dx = 0, \cdots, \int_0^1 kU_{n-1} u dx = 0,$$

the integral $D(u)$ has a minimum value and this value is the constant λ_n described above. Furthermore the minimizing function giving $D(u)$ the value λ_n is $u(x) = \pm U_n(x)$. By an application of some known results from the theory of differential equations and the Jacobi condition of the calculus of variations, it was then proved that U_i vanishes $i-1$ times in the interval $0<x<1$.

Professor G. A. Bliss suggested the consideration of the above problem without recourse to integral equation theory. As a result of his suggestions a method was developed which made use of known theorems in the calculus of variations and which, in some respects, was readily applied to a more general case.

In this paper the boundary value problem is considered which arises from the problem of minimizing the second variation of the problem of the calculus of variations for a space of $n+1$ dimensions. This second variation is

$$I_2(\eta) = \int_{x_1}^{x_2} 2\omega_0(x, \eta, \eta') dx,$$

where

$$2\omega_0(x, \eta, \eta') = P_{ij}\eta_i\eta_j + 2Q_{ij}\eta_i\eta_j' + R_{ij}\eta_i'\eta_j' \qquad (i, j = 1, \cdots, n),$$

and where the elements of η_i form a set of so-called admissible variations. It is understood that in the formula for $2\omega_0$ the summation convention of tensor analysis is used. A subscript i or j appearing twice in the same term indicates that the term is to be summed from 1 to n with respect to that subscript. It is supposed that the quadratic form $R_{ij}\eta_i'\eta_j'$ is positive definite on the interval $x_1 \leqq x \leqq x_2$ and that the coefficients of $2\omega_0$ are continuous functions of x with continuous second derivatives. In addition to the extension of Richardson's results to a space of $n+1$ dimensions, it will be noted that the integral $I_2(\eta)$ is of a less restricted form than the integral treated in his paper.

In §1 the matrix notation is introduced for the sake of brevity and

simplicity of expression, and a few preliminary remarks are made. In §2 we define the successive classes H_i, H_i' $(i = 1, 2, \cdots)$ of arcs $\eta(x)$ in which we are to consider the problem of minimizing $I_2(\eta)$. In §3 it is proved by methods of the calculus of variations that $I_2(\eta)$ has a greatest lower bound λ_0 in the class H_0', and that in the class H_0 there exists a set of r_0 linearly independent arcs for which

$$I_2(\eta) - \lambda_0 \int_{x_1}^{x_2} \eta\eta dx = 0.$$

In §4 it is proved that $I_2(\eta)$ has a minimum value λ_i in the general class H_i', and that in the class H_i there exists a set of r_i linearly independent arcs for which

$$I_2(\eta) - \lambda_i \int_{x_1}^{x_2} \eta\eta dx = 0.$$

The third and fourth sections thus carry through an induction which enables one to arrive at a set of existence theorems for the boundary value problem with a parameter. These results are a generalization of the existence theorem in Richardson's paper. One would not expect to find a direct generalization of his oscillation theorem, but it is very probable that there exist results for a system of ordinary linear differential equations of the second order with a parameter, which would be somewhat analogous to his oscillation theorem.

The theorems from the calculus of variations to which one will refer in §4 are contained in most convenient form in a set of lectures* on the Lagrange problem given by Professor Bliss at Chicago. Analogous theorems for the problem without adjoined differential equations will be used to obtain the results of §3.

In preparing this dissertation I have had the privilege of working under Professor Bliss and I wish to acknowledge his never failing advice and direction.

1. Notations and preliminary remarks

The problem of the calculus of variations is that of finding among the arcs

$$y_i = y_i(x) \qquad (i = 1, \cdots, n\,;\, x_1 \leqq x \leqq x_2) \tag{3}$$

which join two given points 1 and 2 in the $(x, y_1, \cdots, y_n)$-space, one which minimizes an integral of the form

* G. A. Bliss, *The Problem of Lagrange in the Calculus of Variations*, a mimeographed record of a set of lectures given at the University of Chicago in the summer of 1925. (Reference A.) See also Bolza, *Vorlesungen über Variationsrechnung*, pp. 542 ff.

$$I_0 = \int_{x_1}^{x_2} f(x, y, y')dx,$$

where y and y' are symbols for the vectors $(y_1, \cdots, y_n)$ and $(y_1', \cdots, y_n')$ respectively.

The matrix and vector notation will be used. The symbols

$$y, z, u, v, J, \xi, \eta, \zeta, \mu, \phi$$

will always represent vectors. Italic capitals, unless otherwise stated, will symbolize matrices. Any small symbol with a bracketed or double superscript will always represent a vector and any symbol with a subscript will stand for a single element. The meaning of other symbols which may occur will be clear from the context. If, for example, Q is a matrix and η a vector in this simplified notation, then $Q\eta$ represents a vector $(Q_{i1}\eta_1 + \cdots + Q_{in}\eta_n;\ i = 1, \cdots, n)$ and ηQ a vector $(\eta_1 Q_{1i} + \cdots + \eta_n Q_{ni};\ i = 1, \cdots, n)$, which shows that $Q\eta = \eta\bar{Q}$ where $\bar{Q}$ denotes the transpose of Q. We have, in fact, the relations

$$Q\eta\eta' = (Q\eta)\eta' = \eta'(Q\eta) = (\eta'Q)\eta = \bar{Q}\eta'\eta = (\eta\bar{Q})\eta' = \eta(\bar{Q}\eta').$$

The expression $\eta\xi$ will often be used for $I\eta\xi$ where I is the identity matrix.*

The assumption will be made as usual that in a neighborhood N of the values (x, y, y') on the minimizing arc the function f has continuous derivatives up to and including those of the fourth order. An admissible arc is defined to be an arc of the form (3) which is continuous and consists of a finite number of arcs each of which possesses a continuous derivative. A variation arc $\eta(x)$ must be admissible.

The second variation† of the integral I_0 is

$$I_2(\eta) = \int_{x_1}^{x_2} 2\omega_0(x, \eta, \eta')dx$$

where

$$2\omega_0(x, \eta, \eta') = f_{yy}\eta\eta + 2f_{y'y}\eta\eta' + f_{y'y'}\eta'\eta'.$$

Writing $P = f_{yy}$, $Q = f_{y'y}$ and $R = f_{y'y'}$, one has

$$I_2(\eta) = \int_{x_1}^{x_2} (P\eta\eta + 2Q\eta\eta' + R\eta'\eta')dx.$$

---

* For a more detailed explanation of the matrix notation see G. A. Bliss, Annals of Mathematics, (2), vol. 6 (1905), p. 58.

† J. Hadamard, *Leçons sur le Calcul des Variations*, vol. 1 p. 317.

On account of the assumptions made concerning the function f, the matrices P, Q and R are continuous and have continuous derivatives of the first two orders. It will be also assumed that the quadratic form $R\eta\eta$ is positive definite on $x_1 \leqq x \leqq x_2$. Then the determinant of the matrix R will be different from zero on that interval.

If E_{12} is a minimizing arc, then $I_2(\eta)$ must be greater than or equal to zero for every set of admissible variations satisfying the end conditions $\eta(x_1) = \eta(x_2) = 0$. This result suggests a new minimum problem in the $x\eta$-space with the integral I_2 in place of I_0 and the end points $(x, \eta) = (x_1, 0), (x, \eta) = (x_2, 0)$ in place of the end points of E_{12}. By a solution of a system of ordinary linear homogeneous differential equations of the second order will be meant an arc, having continuous first and second derivatives, defined by a vector $\eta(x)$ which satisfies these equations. The class H_0 is defined to be all admissible arcs $\eta(x)$ for which $\eta(x_1) = \eta(x_2) = 0$, and the problem in $x\eta$-space may be stated as that of minimizing $I_2(\eta)$ in the class of arcs H_0. Since $|R| \neq 0$ on $x_1 \leqq x \leqq x_2$ and $|\omega_{0\eta'\eta'}| = |R|$, it follows* that an arc which minimizes $I_2(\eta)$ in the class H_0 must be a solution of the system of equations

$$J(\eta) = 0, \quad \eta(x_1) = \eta(x_2) = 0, \tag{4}$$

where

$$J(\eta) = \frac{d}{dx}\omega_{0\eta'} - \omega_{0\eta} = \frac{d}{dx}(Q\eta + R\eta') - (P\eta + \bar{Q}\eta').$$

The system of equations (4) is linear and homogeneous in the variables η, η', η'' and the determinant of the variables η'' is the determinant of the matrix $\omega_{0\eta'\eta'}$, which is equal to the determinant of R. The determinant $|R|$ is different from zero along E_{12} and hence from the theory of differential equations it follows that the system (4) has one and but one solution η taking prescribed values of η, $\omega_{0\eta'}$ at a given value of x. A matrix of n solutions $u^{(k)}$ $(k = 1, \cdots, n)$ exists therefore for which the matrix $U = \|u^{(k)}\|$ is the identity matrix and the corresponding matrix of the functions $\omega_{0\eta'}$ has all its elements zero at the point 1. Also there exists a matrix of n solutions $v^{(k)}$ $(k = 1, \cdots, n)$ for which the matrix $V = \|v^{(k)}\|$ has all its elements zero at the point 1 and the corresponding matrix of the functions $\omega_{0\eta'}$ is the identity matrix at that point. Furthermore the columns of U and V form $2n$ linearly independent solutions of the system (4). If u and v are two solutions of (4) for which the constant expression

$$u\omega_{0v} - v\omega_{0u}$$

* Hadamard, loc. cit., p. 67, p. 73.

is zero, they are said to be conjugate solutions.* The solutions $u^{(k)}$ $(k=1, \cdots, n)$ are conjugate pairs, since their functions $\omega_{0\eta'}$ all vanish for $x=x_1$.

2. The classes of arcs H_i and H_i'

We will now define the classes of arcs for the various minimum problems to be studied and will sketch, as far as is necessary for that purpose, the methods of the next two sections. The class of arcs H_0 is by definition made up of all arcs $\eta(x)$ which satisfy the conditions $\eta(x_1)=0$, $\eta(x_2)=0$ and which possess continuous derivatives on the interval $x_1 \leqq x \leqq x_2$ except possibly at a finite number of corners, and the class H_0' is the totality of arcs of the class H_0 which satisfy the further condition

$$\int_{x_1}^{x_2} \eta\eta dx = 1.$$

Arcs which satisfy this condition will be called normed. The problem of minimizing $I_2(\eta)$ in the class H_0' will be considered in the next section. It will be proved that $I_2(\eta)$ has a minimum value λ_0 in the class H_0', and that there exists a set of linearly independent arcs η^{0k} $(k=1, \cdots, r_0)$ in the class H_0 such that

$$I_2(\eta) - \lambda_0 \int_{x_1}^{x_2} \eta\eta dx = 0.$$

The functions η^{0k} $(k=1, \cdots, r_0)$ are called the characteristic functions for the class H_0. The class H_1 is now defined to be all those arcs of the class H_0 for which the conditions

$$\int_{x_1}^{x_2} \eta^{0k}\eta dx = 0 \qquad (k = 1, \cdots, r_0)$$

are satisfied. The class H_1' is then taken to be all normed arcs of the class H_1. By methods developed in §3 it is possible to prove directly that there exists a set of linearly independent characteristic functions η^{1k} $(k=1, \cdots, r_1)$ in the class H_1 for which

$$I_2(\eta) - \lambda_1 \int_{x_1}^{x_2} \eta\eta dx = 0,$$

where λ_1 is the greatest lower bound of the values of $I_2(\eta)$ in the class H_1'. The class H_2 is then defined to be all arcs of H_1 for which

$$\int_{x_1}^{x_2} \eta^{1k}\eta dx = 0 \qquad (k = 1, \cdots, r_1),$$

and the class H_2' consists of all normed arcs of the class H_2.

* Bliss, loc. cit. A, p. 73; von Escherich, Wiener Berichte, vol. 107, p. 1244.

Supposing the existence of a set of characteristic functions in the class H_2, one defines the classes H_3 and H_3' in a manner similar to the definitions of the classes H_2 and H_2'. One proceeds in this manner and defines the general classes H_i and H_i', the class H_i being all those arcs in the class H_{i-1} for which

$$\int_{x_1}^{x_2} \eta^{i-1,k}\eta dx = 0 \qquad (k = 1, \cdots, r_{i-1})$$

where, if λ_{i-1} is the minimum value of $I_2(\eta)$ in the class H_{i-1}', the $\eta^{i-1,k}$ are the r_{i-1} linearly independent characteristic functions in the class H_{i-1} for which

$$I_2(\eta) - \lambda_{i-1}\int_{x_1}^{x_2} \eta\eta dx = 0.$$

The class H_i' consists of all normed arcs of the class H_i. That is, the class H_i' is made up of all arcs η which have continuous derivatives except at a finite number of points on the interval $x_1 \leqq x \leqq x_2$ and which satisfy the conditions

$$\eta(x_1) = \eta(x_2) = 0, \qquad \int_{x_1}^{x_2} \eta\eta dx = 1,$$

$$\int_{x_1}^{x_2} \eta^{0,k}\eta dx = 0 \ (k = 1, \cdots, r_0), \cdots, \int_{x_1}^{x_2} \eta^{i-1,k}\eta dx = 0 \ (k = 1, \cdots, r_{i-1}).$$

3. The minimum of the second variation in the class H_0'

The problem of minimizing $I_2(\eta)$ in the class H_0' is that of finding among all admissible arcs η satisfying the conditions

$$\int_{x_1}^{x_2} \eta\eta dx = 1, \quad \eta(x_1) = \eta(x_2) = 0$$

one for which the integral $I_2(\eta)$ takes a minimum value. If one introduces the new variable

$$\zeta_0 = \int_{x_1}^{x_2} \eta\eta dx$$

the problem becomes that of finding among the arcs $\eta(x)$, $\zeta_0(x)$ joining the two points $(x, \eta, \zeta_0) = (x_1, 0, 0)$ and $(x, \eta, \zeta_0) = (x_2, 0, 1)$ and satisfying the equation

$$\phi_0 \equiv \zeta_0' - \eta\eta = 0$$

one which minimizes the integral

$$I_2(\eta) = \int_{x_1}^{x_2} 2\omega_0(x, \eta, \eta')dx.$$

That is, the isoperimetric problem has been replaced by an equivalent Lagrange problem of the calculus of variations. Having set up the form

$$\Omega(x,\eta,\eta') = 2\omega_0 + \lambda(\zeta_0' - \eta\eta)$$

one notes that $|\Omega_{\eta'\eta'}|$ is different from zero on $x_1 \leqq x \leqq x_2$. The first necessary condition* of the Lagrange problem gives then the following result:

THEOREM 1. *An arc $\eta(x)$ which minimizes $I_2(\eta)$ in the class H_0' must be a solution of the boundary value problem*

$$J(\eta) + \lambda\eta = 0, \qquad \eta(x_1) = \eta(x_2) = 0.$$

It is further possible to prove

THEOREM 2. *The values of the integral $I_2(\eta)$ in the class H_0' have a greatest lower bound λ_0.*

To prove this a new vector ζ is introduced by means of the n equations

$$\zeta = \omega_{0\eta'} = Q\eta + R\eta'.$$

Since the determinant of R is different from zero on $x_1 \leqq x \leqq x_2$, these equations may be solved for the elements of the vector η' in the form

$$\eta' = R^{-1}\zeta - R^{-1}Q\eta = \zeta R^{-1} - \eta\bar{Q}R^{-1}.$$

One substitutes this expression for η' in the integral

$$I_2(\eta) - \lambda\int_{x_1}^{x_2}\eta\eta dx = \int_{x_1}^{x_2}[(P - \lambda I)\eta\eta + 2Q\eta\eta' + R\eta'\eta']dx$$

and finds that the integral takes the form

$$\int_{x_1}^{x_2}[(A - \lambda I)\eta\eta + R^{-1}\zeta\zeta]dx,$$

where $A = P - \bar{Q}R^{-1}Q$.

The matrix of the quadratic form $(A - \lambda I)\eta\eta$ is

$$A - \lambda I = \begin{Vmatrix} a_{11} - \lambda & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} - \lambda & \cdots & a_{2n} \\ \cdot & \cdot & \cdots & \cdot \\ a_{n1} & \cdot & \cdots & a_{nn} - \lambda \end{Vmatrix}$$

where the a_{ij} are independent of λ. The chain of principal minors from $a_{11} - \lambda$ to the determinant of the matrix $A - \lambda I$ have respectively $-\lambda$, λ^2, $\cdots$, $(-1)^n\lambda^n$ as the dominant terms in λ. Certainly for λ sufficiently large and

* Bliss, loc. cit. A, p. 10.

negative these determinants will all be positive.* That is, for a sufficiently large negative value of λ the form $(A-\lambda I)\eta\eta$ will be positive definite for all x on $x_1 \leqq x \leqq x_2$. Also under the transformation $u = R^{-1}\zeta\zeta$ the form Ruu becomes $R^{-1}\zeta\zeta$. Hence the assumption that Ruu is positive definite implies the same property for $R^{-1}\zeta\zeta$. There exists therefore a smallest value λ_0 of λ such that

$$I_2(\eta) - \lambda_0 \int_{x_1}^{x_2} \eta\eta dx \geqq 0$$

for every η in the class H_0. Hence the theorem.

The following lemma is of importance in the proof of later theorems.

LEMMA 1. *Along every admissible arc η which is a solution of the boundary value problem*

$$J(\eta) + \lambda\eta = 0, \quad \eta(x_1) = \eta(x_2) = 0 \tag{5}$$

the integral

$$I_2(\eta) - \lambda \int_{x_1}^{x_2} \eta\eta dx$$

has the value zero.

For the proof one makes use of a well known property of quadratic forms to write the integral $I_2(\eta)$ as

$$I_2(\eta) = \int_{x_1}^{x_2} 2\omega_0 dx = \int_{x_1}^{x_2} (\eta\omega_{0\eta} + \eta'\omega_{0\eta'})dx.$$

On integrating the second term of the right member by parts, this becomes

$$I_2(\eta) = \eta\omega_{0\eta'}\Big|_{x_1}^{x_2} + \int_{x_1}^{x_2} \eta\left(\omega_{0\eta} - \frac{d}{dx}\omega_{0\eta'}\right)dx.$$

If η is a solution of the boundary value problem (5), one notes immediately that

$$I_2(\eta) = \lambda \int_{x_1}^{x_2} \eta\eta dx.$$

COROLLARY. *Along every normed admissible arc which is a solution of the boundary value problem* (5) *the integral $I_2(\eta)$ has the value λ.*

---

* It is of interest to note that for a sufficiently large and positive λ the principal minor determinants of the chain will alternate in sign. If $R\eta\eta$ is supposed negative definite on $x_1 \leqq x \leqq x_2$, then the values of $I_2(\eta)$ have a least upper bound in the class H_0'.

The problem of minimizing the integral

$$I_2(\eta) - \lambda_0 \int_{x_1}^{x_2} \eta\eta dx \tag{6}$$

in the class H_0 will now be considered. The arc $\eta = 0$ is a minimizing arc. For if there existed an arc u in H_0 for which the integral (6) were less than zero, then the vector

$$\eta = u\left[\int_{x_1}^{x_2} uudx\right]^{-1/2}$$

would be in the class H_0' and would give $I_2(\eta)$ a value less than λ_0 which is a contradiction.

The second variation of (6) is

$$\int_{x_1}^{x_2} (\Omega_{0\eta\eta}\xi\xi + 2\Omega_{0\eta'\eta}\xi\xi' + \Omega_{0\eta'\eta'}\xi'\xi')dx, \tag{7}$$

where

$$2\Omega_0(x,\eta,\eta') = (P - \lambda_0 I)\eta\eta + 2Q\eta\eta' + R\eta'\eta'$$

and ξ is a set of admissible variations. The Euler system and the Jacobi* system of n differential equations for the problem of minimizing the integral (6) are therefore identical; namely

$$J(\eta) + \lambda\eta = 0 \tag{8}$$

for $\lambda = \lambda_0$.

THEOREM 3. *There exists a set of r_0 linearly independent arcs $\eta^{0k}(x)$ $(k = 1, \cdots, r_0)$ in the class H_0 such that*

$$I_2(\eta^{0k}) - \lambda_0 \int_{x_1}^{x_2} \eta^{0k}\eta^{0k}dx = 0 \qquad (k = 1, \cdots, r_0).$$

For the proof one considers the problem of minimizing the integral (6) in the class H_0. As noted above, the Euler and the Jacobi differential equations for this problem are identical. In §1 the solutions of such a system were considered and a matrix of solutions V was uniquely determined whose elements all vanish at $x = x_1$ and such that the matrix of the corresponding functions $\omega_{0\eta'}$ was the identity matrix I. Also if λ is allowed to vary the solutions of the system (8) are continuous functions of λ.† The determinant

* Bliss, loc. cit. A, pp. 57–63. The minimum problem discussed above has no adjoined differential equations, and theorems, corresponding to those of the Lagrange problem, hold for this more special case.

† J. Horn, *Gewöhnliche Differentialgleichungen beliebiger Ordnung*, p. 293.

of the matrix V is a Mayer* determinant $\Delta(x, x_1, \lambda_0)$ for our minimizing problem, since its columns are n linearly independent solutions of the Jacobi differential equations and its elements vanish at $x = x_1$. The vanishing of $\Delta(x, x_1, \lambda_0)$ for a value of x between x_1 and x_2 implies that there exists an admissible arc such that the second variation (7) is less than zero.* Since (7) is the same as the integral (6), this result contradicts the previous statement that λ_0 is the minimum value of $I_2(\eta)$ in the class H_0'. Hence $\Delta(x, x_1, \lambda_0)$ cannot vanish for a value x between x_1 and x_2.

In order to prove that $\Delta(x, x_1, \lambda_0)$ must vanish at $x = x_2$, it will now be assumed that $\Delta(x, x_1, \lambda_0)$ is different from zero on $x_1 < x \leqq x_2$ and it will then be shown that this leads to a contradiction. The determinant of the matrix U with initial value I at $x = x_1$, as set up in §1, is a continuous function of λ which is different from zero at $x = x_1$. Also the columns of U form a conjugate system of solutions of the Jacobi differential equations. If λ is allowed to vary, the matrix U with the same initial conditions as before will still have for columns a conjugate system of solutions of the corresponding Jacobi differential equations, and for h sufficiently small and positive there will be an interval $x_1 - h \leqq x \leqq x_1 + h$ on which $U(x, \lambda_0 + \epsilon)$ will be different from zero. It is known† that $\Delta(x, x_1, \lambda_0 + \epsilon)$ cannot vanish for a pair of points x', x'' on an interval $x_1 - h \leqq x \leqq x_1 + h$ for which there is a conjugate system with determinant $|U(x, \lambda_0 + \epsilon)| \neq 0$. Thus $\Delta(x, x_1, \lambda_0 + \epsilon)$ is different from zero on $x_1 < x \leqq x_1 + h$. From continuity considerations $\Delta(x, x_1, \lambda_0 + \epsilon)$ is different from zero on $x_1 + h \leqq x \leqq x_2$ if also $\Delta(x, x_1, \lambda_0)$ is different from zero on that interval, ϵ being positive and sufficiently small. That is, $\Delta(x, x_1, \lambda_0) \neq 0$ on $x_1 < x \leqq x_2$ implies $\Delta(x, x_1, \lambda_0 + \epsilon) \neq 0$ on $x_1 < x \leqq x_2$ for ϵ positive and sufficiently small.

Consider now the problem of minimizing

$$I_2(\eta) - (\lambda_0 + \epsilon) \int_{x_1}^{x_2} \eta\eta dx \tag{9}$$

in the class H_0. If $\Delta(x, x_1, \lambda_0 + \epsilon)$ is different from zero on the interval $x_1 < x \leqq x_2$, there exists no conjugate point‡ on this interval for the extremal arc $\eta = 0$. The Clebsch§ quadratic form is $R\pi\pi$ which is greater than zero for every set $\pi = (\pi_1, \cdots, \pi_n) \neq (0, \cdots, 0)$. The Weierstrass§ function $E(x, \eta, \eta', \zeta')$ is found to be

$$R(\zeta'\zeta' - 2\eta'\zeta' + \eta'\eta') = R(\zeta' - \eta')(\zeta' - \eta')$$

* See first footnote on page 572.

† Bliss, loc. cit. A, p. 75; Bolza, loc. cit., p. 632.

‡ Bliss, loc. cit. A, p. 52; Bolza, loc. cit., p. 610.

§ Bliss, loc. cit. A, pp. 49–51.

which is greater than zero for every $\zeta' \neq \eta'$. From the sufficiency theorem* of the calculus of variations it therefore follows that the extremal arc $\eta = 0$, containing no conjugate point on $x_1 < x \leqq x_2$, gives (9) a smaller value than any other arc ξ of the class H_0 which is not identically zero. If then $\Delta(x, x_1, \lambda_0)$ is different from zero on $x_1 < x \leqq x_2$, one finds that $\Delta(x, x_1, \lambda_0 + \epsilon)$ has the same property and hence $I_2(\eta) > \lambda_0 + \epsilon$ for every arc of the class H_0'. This contradicts the hypothesis that λ_0 is the greatest lower bound of $I_2(\eta)$ in the class H_0'. Therefore $\Delta(x_2, x_1, \lambda_0) = 0$. That is, the columns of the matrix V are linearly dependent at $x = x_2$.

If the rank of the matrix V is $n - r_0$ at $x = x_2$, the equations

$$V(x_2)c = 0,$$

where $c = (c_1, \cdots, c_n)$ is independent of x, have r_0 linearly independent solutions $c \neq 0$ and these define a fundamental system of solutions of these linear equations in terms of which all other solutions are linearly expressible. Moreover for each such c the vector $\eta = Vc$ is a solution of the boundary value problem (5) corresponding to the value $\lambda = \lambda_0$. Hence by Lemma 1 there exist r_0 linearly independent arcs $\eta^{0k} (k = 1, \cdots, r_0)$ of the class H_0 such that

$$I_2(\eta) - \lambda_0 \int_{x_1}^{x_2} \eta\eta dx = 0.$$

COROLLARY. *There exists a set of r_0 linearly independent arcs in the class H_0' for which the integral $I_2(\eta)$ takes the value λ_0.*

4. THE MINIMUM OF THE SECOND VARIATION IN THE CLASS H_i'

In the previous section it has been proved that $I_2(\eta)$ has a minimum value λ_0 in the class H_0', and that there exists a set of r_0 linearly independent arcs in the class H_0 for which the relation

$$I_2(\eta) - \lambda_0 \int_{x_1}^{x_2} \eta\eta dx = 0 \tag{10}$$

holds. The classes H_1 and H_1' will then be defined as in §2. Since H_1' is a sub-class of H_0' it follows that $I_2(\eta)$ has a greatest lower bound λ_1 in the class H_1' and that λ_1 cannot be less than λ_0. In order to prove by induction that there exists in the general class H_i a set of r_i linearly independent arcs for which the integral

$$I_2(\eta) - \lambda_i \int_{x_1}^{x_2} \eta\eta dx \tag{11}$$

* Bliss, loc. cit. A, p. 70; Bolza, loc. cit., p. 638.

is zero, we will assume the existence of such a set for each of the classes H_1, $H_2, \cdots, H_{i-1}$ with λ_i replaced by $\lambda_1, \lambda_2, \cdots, \lambda_{i-1}$, respectively, in the integral (11). We will denote these sets in the classes $H_1, \cdots, H_{i-1}$ by $\eta^{1k}, \cdots, \eta^{i-1,\,k}$ where the range of k will be determined by the first superscript. For example, in the symbol η^{jk}, k has the range $1, \cdots, r_j$. The classes H_i and H_i' will then be defined as in §2. Since H_i' is a sub-class of H_{i-1}' it follows that $I_2(\eta)$ has a greatest lower bound λ_i in the class H_i' and that λ_i cannot be less than λ_{i-1}.

The problem of minimizing the integral (11) in the class H_i will now be considered. If we introduce the $r_0+r_1+\cdots+r_{i-1}$ new variables

$$\zeta_{0k} = \int_{x_1}^{x} \eta^{0k}\eta dx, \quad \zeta_{1k} = \int_{x_1}^{x} \eta^{1k}\eta dx, \cdots, \zeta_{i-1,k} = \int_{x_1}^{x} \eta^{i-1,\,k}\eta dx,$$

set $\rho = r_0+r_1+\cdots+r_{i-1}$, write η^k and $\zeta_k (k = 1, 2, \cdots, \rho)$ as new notations for the sets $\eta^{0k}, \eta^{1k}, \cdots, \eta^{i-1,k}$ and $\zeta_{0k}, \zeta_{1k}, \cdots, \zeta_{i-1,k}$, respectively, and let z represent the vector whose elements are η and the ζ_k, the problem becomes that of finding among all the admissible arcs $z(x)$ joining the two points $(x, z) = (x_1, 0)$ and $(x, z) = (x_2, 0)$, and satisfying the equations

$$\phi_k \equiv \zeta_k' - \eta^k\eta = 0,$$

one which gives the integral (11) a minimum value. The problem is therefore a Lagrange problem with ρ adjoined differential equations. The arc $z = 0$ is a minimizing arc. For, if there existed a vector u in the class H_i for which the integral (11) were less than zero, then the arc

$$\eta = u\left(\int_{x_1}^{x_2} uudx\right)^{-1/2}$$

would be in the class H_i' and would give $I_2(\eta)$ a value less than λ_i, which is a contradiction.

Writing μ and ϕ as symbols for the vectors $(\mu_1, \cdots, \mu_\rho)$ and $(\phi_1, \cdots, \phi_\rho)$, one sets up the form

$$2\omega_i = 2\omega_0 - \lambda_i\eta\eta + 2\mu\phi,$$

where the elements of μ are Lagrange multipliers to be determined. The determinant

$$R_i \equiv \begin{vmatrix} \omega_{iz'z'} & \phi_{z'} \\ \phi_{z'} & 0 \end{vmatrix}$$

is readily found to equal $-|R|$. Hence R_i is different from zero on $x_1 \leqq x \leqq x_2$.

Therefore it follows* that an arc which minimizes the integral (11) in the class H_i must be a solution of the Euler-Lagrange system of equations

$$\frac{d}{dx}\omega_{iz'} - \omega_{iz} = 0, \quad \phi = 0.$$

These equations are the system

$$J(\eta) + \lambda_i\eta + \mu_\beta\eta^\beta = 0, \quad \phi = 0 \qquad \left(\mu_\beta\eta^\beta \equiv \sum_{\beta=1}^{\rho}\mu_\beta\eta^\beta\right), \tag{12}$$

with the equations $d\mu_\beta/dx = 0$ which say that the μ_β are constants.

For a quadratic form $2\omega_i$ in the integrand it is readily seen that the second variation is a quadratic form with the same coefficients. The equations of variation† of the equations $\phi=0$ are the same equations since the ϕ are linear. The equations (12) are linear. It follows readily therefore that the accessory equations‡ for the minimum problem under consideration are the Euler-Lagrange equations (12) themselves.

In the proof of the existence theorem of this section use will be made of the two following lemmas:

LEMMA 2.§ *The sets η^k are linearly independent.*

For, suppose $c_\beta\eta^\beta = 0$ $(\beta = 1, \cdots, \rho)$. Then, for every η^k,

$$0 = \int_{x_1}^{x_2} c_\beta\eta^\beta\eta^k dx = \int_{x_1}^{x_2} c_\gamma\eta^\gamma\eta^k dx,$$

where γ ranges over the indices for which η^k, η^γ are in the same set $\eta^{ij}(j = 1, \cdots, r_i)$. Therefore

$$\int_{x_1}^{x_2} c_\gamma\eta^\gamma c_\delta\eta^\delta dx = 0,$$

the range of δ being the same as that of γ. That is, $c_\gamma\eta^\gamma$ is identically zero and hence $c_\gamma = 0$. A similar argument shows that all of the other coefficients c_β are zero.

---

* Bliss, loc. cit. A, p. 11; Bolza, loc. cit., p. 558.

† Bliss, loc. cit. A, p. 4.

‡ Bliss, loc. cit. A, p. 57; von Escherich, loc. cit., vol. 107, p. 1236.

§ For the proof of Lemmas 2 and 3 it is convenient to use the notation of tensor analysis, along with the matrix notation. For example,

$$c_\beta\eta^\beta = \sum_{\beta=1}^{\rho} c_\beta\eta^\beta$$

stands for the vector $(c_1\eta_1{}^1 + \cdots + c_\rho\eta_1{}^\rho, \cdots, c_1\eta_n{}^1 + \cdots + c_\rho\eta_n{}^\rho)$.

LEMMA 3. *The Lagrange multipliers μ_β are zero for every solution $\eta(x)$ of* (12) *which satisfies the conditions*

$$\eta(x_1) = \eta(x_2) = 0, \quad \zeta_k(x_2) = \int_{x_1}^{x_2} \eta^k \eta dx = 0.$$

From the definition of ζ_k it follows also that $\zeta_k(x_1) = 0$. For the proof of the lemma we have

$$J(\eta) + \lambda_i \eta + \mu_\beta \eta^\beta = 0, \; J(\eta^k) + \lambda' \eta^k = 0, \; \eta^k(x_1) = \eta^k(x_2) = 0,$$

where λ' is one of the constants $\lambda_1, \lambda_2, \cdots, \lambda_{i-1}$. Hence

$$\eta^k J(\eta) - \eta J(\eta^k) + (\lambda_i - \lambda')\eta^k \eta + \mu_\beta \eta^\beta \eta^k = 0,$$

or, after some computation,

$$\frac{d}{dx}(\eta^k \omega_{0\eta'} - \eta \omega_{0\eta^{k'}}) + (\lambda_i - \lambda')\eta^k \eta + \mu_\beta \eta^\beta \eta^k = 0.$$

Integrating from x_1 to x_2 and using the hypotheses of the lemma, one obtains

$$\mu_\beta \int_{x_1}^{x_2} \eta^\beta \eta^k dx = 0 \qquad (\beta, k = 1, \cdots, \rho).$$

If the μ_β are not all zero, we find

$$\int_{x_1}^{x_2} \mu_\beta \eta^\beta \mu_\gamma \eta^\gamma dx = 0 \qquad (\beta, \gamma = 1, \cdots, \rho).$$

Therefore $\mu_\beta \eta^\beta$ is identically zero and hence by the preceding lemma $\mu_\beta = 0$.

THEOREM 4. *There exists a set of r_i linearly independent arcs $\eta^{ik}(x)$ $(k = 1, \cdots, r_i)$ in the class H_i such that*

$$I_2(\eta^{ik}) - \lambda_i \int_{x_1}^{x_2} \eta^{ik} \eta^{ik} dx = 0 \qquad (k = 1, \cdots, r_i).$$

Moreover λ_i is definitely greater than λ_{i-1}.

For the proof one considers the problem of minimizing the integral (11) in the class H_i. As noted above, the Euler-Lagrange and the accessory system of differential equations are identical, namely

$$J(\eta) + \lambda \eta + \mu_\beta \eta^\beta = 0, \qquad \phi = 0, \tag{13}$$

for $\lambda = \lambda_i$. This system is linear and homogeneous in the variables z, z', z'', μ_β, μ_β' and the determinant of the variables z'', μ_β' is equal to the determinant of the matrix R. In §1, the solutions of such a system were considered

and it was noted that a matrix V can be uniquely determined whose columns are $n+\rho$ linearly independent solutions $z=(\eta, \zeta)$ of (13), whose elements all vanish at $x=x_1$, and such that the matrix of the $n+\rho$ corresponding fun ctions ω_{ir} is the identity matrix I at $x=x_1$. Also if λ is allowed to vary, the solutions of the system (13) are continuous functions of λ. The determinanteof the matrix V is a Mayer determinant* for our minimizing problem, since its columns are $n+\rho$ linearly independent solutions of the accessory syst m which vanish at $x=x_1$. The vanishing of $|V| = \Delta(x,x_1,\lambda_i)$ for a value of $x$between x_1 and x_2 implies† that there exists an admissible set ξ of $n+\rho$ functions which substituted for η, ζ make the second variation of our minimum problem less than zero. Since this second variation is the same as the integral (11), we have a contradiction of the previous statement that λ_i is the minimum value of $I_2(\eta)$ in the class H_i'. Hence $\Delta(x, x_1, \lambda_i)$ cannot vanish for a value of x between x_1 and x_2.

The proof that $\Delta(x, x_1, \lambda_i)$ must vanish at $x=x_2$ follows by an argument corresponding to that of Theorem 3, if λ_0 is replaced by λ_i, $\Delta(x, x_1, \lambda_0)$ by $\Delta(x, x_1, \lambda_i)$ and the matrix U by its corresponding matrix. Since $\Delta(x_2, x_1, \lambda_i)=0$, it follows that the columns of the matrix V are linearly dependent at $x=x_2$.

If the rank of the matrix V is $n+\rho-r_i$ at $x=x_2$, the equations

$$V(x_2)c = 0,$$

where $c=(c_1, \cdots, c_{n+\rho})$ is independent of x, have r_i linearly independent solutions $c\neq 0$, which define a fundamental system of solutions of these linear equations in terms of which all other solutions different from zero are linearly expressible. Moreover, for each such c the vector $w=Vc$ is a solution of the system of equations (12) with elements all vanishing at x_1 and x_2. The conditions of Lemma 3 are satisfied. Hence by Lemma 1 there exist r_i linearly independent arcs

$$\eta^{ik}, \zeta_1{}^{ik}, \cdots, \zeta_\rho{}^{ik} \qquad (k = 1, \cdots, r_i)$$

of the class H_i for which the η^{ik} satisfy equation (10) with λ_0 replaced by λ_i.

To prove $\lambda_i > \lambda_{i-1}$, it is first shown that no η^{ik} can be a linear combination of the $\eta^{i-1,k}$. This follows from the proof of Lemma 2 if we let the sets η^k include the η^{ij} as well as $\eta^{0j}, \cdots, \eta^{i-1,j}$. In the class H_{i-1} the only sets η making

* Bliss, loc. cit. A, p. 63; Bolza, loc. cit., p. 611.

† Bliss, loc. cit. A, p. 59; Bolza, loc. cit., p. 617.

$$I_2(\eta) - \lambda_{i-1} \int_{x_1}^{x_2} \eta\eta dx = 0$$

are the sets $\eta^{i-1,k}$ $(k = 1, \cdots, r_{i-1})$ and linear combinations of them, as was found in the preceding pages. As just shown above, no η^{ik} can be such a linear combination. But η^{ik} satisfies the relation

$$I_2(\eta) - \lambda_i \int_{x_1}^{x_2} \eta\eta dx = 0.$$

Hence $\lambda_i \neq \lambda_{i-1}$, and since λ_i cannot be less than λ_{i-1} one has $\lambda_i > \lambda_{i-1}$. This completes the proof of Theorem 4, and one readily sees that the following corollary is also true:

COROLLARY. *There exists a set of r_i linearly independent arcs in the class H_i' for which the integral $I_2(\eta)$ takes the value λ_i.*

BROWN UNIVERSITY,
PROVIDENCE, R. I.

AN EXTENSION OF PASCAL'S THEOREM*

BY

CHARLES A. RUPP

INTRODUCTION

In 1825 the Académie Royale de Bruxelles proposed as a prize topic the extension of Pascal's theorem to space of three dimensions. The prize was won by Dandelin,† who showed that a skew hexagon formed of three lines from each regulus of a hyperboloid of revolution has the Pascal property that pairs of opposite planes meet on a plane, and the dual, or Brianchon property, that the lines joining pairs of opposite vertices meet in a point. Hesse‡ wrote several papers on the Dandelin skew hexagons, emphasizing the polar properties of the Pascal plane and the Brianchon point with respect to the quadric bearing the two reguli. Several of the older analytic geometries of three dimensions devote some space to the problem of the skew hexagon, as Salmon, and more notably Plücker§, who offered much original material. In a recent article, the present writer‖ has discussed the skew hexagon from an elementary analytic approach.

The foregoing citations exhibit the idea that Pascal's theorem deals with six elements of a quadratic curve, and that the space extension offered will deal with six elements of a quadric surface. It happens that an extension of this sort is valid in space of three dimensions; but in S_n, where $n>3$, it is clear that a skew hexagon of six rulings of a hyperquadric must lie in an S_3 if it is to possess the Brianchon property. It seems, therefore, that the hexagon idea must be abandoned in the search for a valid extension of Pascal's theorem in S_n.

In the plane, a variant of Pascal's theorem affirms that if two triangles are in homology, then the six points of intersection of the sides of the one with the non-corresponding sides of the other lie upon a conic. From a consideration of the converses of this theorem and its dual it occurred in-

* Presented to the Society, April 6, 1928; received by the editors in October, 1928, and (revised) in April, 1929.

† Dandelin, *Mémoire sur l'hyperboloide de révolution et sur les hexagones de Pascal et de M. Brianchon*, Nouvelles Mémoires de l'Académie Royale de Bruxelles, vol. 3 (1826), pp. 1–14.

‡ Hesse, Werke, pp. 58, 651, 676.

§ Plücker, *System der Geometrie des Räumes*, §§87–92.

‖ Rupp, *Stereographic projection of a quadric*, American Mathematical Monthly, vol. 25 (1928), pp. 415–421.

dependently to Chasles* and to Weddle† that the Pascal configuration might conveniently be considered as a property of a pair of triangles whose sides meet on a conic, and that the space extension would concern a pair of tetrahedra and a quadric; in the plane opposite sides of the triangles meet in a triple of points on a line, and hence these geometers reasoned that in space opposite faces of the tetrahedra would meet in a quadruple of lines on a regulus. They readily devised synthetic proofs of the theorem. Chasles made the pregnant observation that the twelve points common to the edges of a given tetrahedron and quadric could be arranged in several ways to define a second tetrahedron which would be effective in the theorem. Weddle went on to study various properties possessed by a pair of effective tetrahedra, and thereby discovered several properties later used by Schläfli in discussing Schläfli simplexes.

Closely allied to the theorem generalized by Chasles and by Weddle is another stating that if two triangles are polar reciprocal with respect to a conic, they are perspective, and hence the lines joining non-corresponding vertices touch a conic, and dually. Schläfli‡ discussed the analogous situation in S_n. His most important discovery was that two simplexes, or complete $(n+1)$-points, of S_n which are polar reciprocal with respect to a hyperquadric have what is now called the Schläfli property, i.e., the $n+1$ S_{n-2}'s of intersection of corresponding S_{n-1}'s (hyperplanes, faces) of the two simplexes lie in such a position that the lines of S_n which meet n of the S_{n-2}'s meet also the other. Following the suggestion of Berzolari,§ one now speaks of Schläfli simplexes, of a Schläfli set of lines or of S_{n-2}'s, or of lines or S_{n-2}'s in the position of Schläfli. Other contributions to the knowledge of the Schläfli situation are due to Brusotti‖ and the present writer.¶ The extension of Pascal's theorem which this paper presents has a close connection with the Schläfli situation.

It will be shown that it is possible to construct, from the points common to a hyperquadric and the edges of a simplex, a certain number of auxiliary simplexes, each of which may be paired with the given simplex to make a Schläfli pair; the ∞^{n-2} lines which meet the S_{n-2}'s of a Schläfli set thus

* Chasles, *Aperçu Historique*, Note 32.

† Weddle, *On theorems in space analogous to those of Pascal and Brianchon in a plane*, Cambridge and Dublin Mathematical Journal, vol. 6 (1851), pp. 116–140.

‡ Schläfli, *Erweiterung des Sätzes, dass zwei polare Dreiecke perspektivisch liegen, auf eine beliebige Zahl von Dimensionen*, Journal für Mathematik, vol. 65 (1866), pp. 189–197.

§ Berzolari, *Sistemi di rette in posizione di Schläfli*, Rendiconti del Circolo Matematico di Palermo, vol. 20 (1905), pp. 229–248.

‖ Brusotti, Rendiconti del Circolo Matematico di Palermo, vol. 20 (1905), pp. 248–261.

¶ Rupp, (unpublished) Chicago (1928) doctoral thesis.

defined lie upon a variety of order and dimension $n-1$, denoted by the symbol V_{n-1}^{n-1}. On each V_{n-1}^{n-1} there are $(n-1)!$ families of generators; since in S_3 the families are called reguli, we shall use the term hyperregulus to refer to a set of ∞^{n-2} lines which are rulings of a V_{n-1}^{n-1}. The older literature about the variety appears to be scanty; Segre* twice mentions it casually in the Encyclopedia. Three papers on the variety have recently been published.†

I. The notation

The figure in S_n consisting of $n+1$ points, which do not lie in the same S_{n-1}, the $(n+1)n/2$ lines joining the points in pairs, the $(n+1)n(n-1)/6$ planes joining the points in triples, $\cdots$, and the $n+1$ S_{n-1}'s joining the points in n-tuples, is called a simplex. The points, lines, and hyperplanes (S_{n-1}'s) of a simplex are called its vertices, edges, and faces, respectively. Choose a given simplex F as the basis of a homogeneous coördinate system; the coördinates of the ith vertex P_i, where i runs from 0 to n, are all zero save at the ith place. The face of F that does not contain P_i is called the face opposite to P_i, and will be denoted by ρ_i; its equation is $x_i=0$.

A hypersurface (variety of dimension $n-1$) of the second order is the locus of points satisfying a quadratic equation; such a hypersurface is called a hyperquadric. Let there be given a hyperquadric of equation

$$Q: \sum a_{ij}x_ix_j = 0 \qquad (i,j = 0,1,\cdots,n; a_{ij} = a_{ji}). \tag{1}$$

The edges of the fundamental simplex F meet the given hyperquadric Q in $2m$ piercing points P_{ij}, where $2m=n(n+1)$, and the points P_{ij}, P_{ji} lie on the line P_iP_j. To find the coördinates of these piercing points, solve the binary quadratic

$$a_{ii}x_i^2 + 2a_{ij}x_ix_j + a_{jj}x_j^2 = 0. \tag{2}$$

It will be convenient to use the quantities defined by

$$\begin{aligned} \Delta_{ij}^2 &= a_{ij}^2 - a_{ii}a_{jj}, \\ b_{ii} &= a_{ii}, \quad b_{ij} = a_{ij} + \Delta_{ij} = b_{ji}, \\ c_{ii} &= a_{ii}, \quad c_{ij} = a_{ij} - \Delta_{ij} = c_{ji}, \end{aligned} \tag{3}$$

in exhibiting the coördinates of P_{ij} and P_{ji} in the form

---

* Segre, Encycklopädie der Mathematischen Wissenschaften, vol. III C7, p. 815 and p. 832.

† Wong, *On a certain system of ∞^{r-2} lines in r-space*, and *On the loci of the lines incident with k $(r-2)$-spaces in S_r*, Bulletin of the American Mathematical Society, vol. 34 (1928), pp. 553–554, and pp. 715–717.

Rupp, *The equation of the V_{n-1}^{n-1} in S_n*, Bulletin of the American Mathematical Society, vol. 35 (1929), pp. 319–320.

(4)

	x_i	x_j	all other coördinates
P_{ij}	b_{ij}	$-b_{ii}$	0
P_{ij}	$-c_{jj}$	c_{ij}	0
P_{ji}	$-b_{jj}$	b_{ij}	0
P_{ji}	c_{ij}	$-c_{ii}$	0 .

Note that the coördinates of either piercing point on an edge of F may be expressed either in terms of b_{ij} or c_{ij}. In handling the first effective simplex which we shall set up, we shall use only the quantities b_{ij}; we shall find that the other effective simplexes can be found from the first, or standard one, by interchanging the elements in certain pairs of the points P_{ij} and P_{ji}. The corresponding change in the analytic work is accomplished by replacing the appropriate b_{ij} by c_{ij}. If the hyperquadric Q is a general one, the quantities b_{ij} and c_{ij} will all be different from zero.

The standard effective grouping of the $2m$ piercing points will be defined by the symbolic matrix

$$G_1 = (P_{ij}) \qquad (i \neq j). \tag{5}$$

The elements of the matrix (P_{ij}) are the piercing points P_{ij}; there are $n+1$ rows and columns, each containing n points. The points of the ith row share the first subscript i, and no pair of them share the second subscript. Geometrically, the points of the ith row lie one on each of the n edges of F which pass through the vertex P_i. Through the points of the ith row can be passed a unique S_{n-1}, which we call π_i and make correspond to the hyperplane ρ_i of F, which is the face of F opposite P_i.

The standard grouping G_1 accordingly defines $n+1$ hyperplanes π_i which constitute the faces of a simplex, T_1, said to be auxiliary to F, the fundamental simplex. The vertices of T_1 may be called the points R_i, R_i being opposite to π_i. The points R_i and P_i are said to be corresponding points.

Consider the problem of determining all the groups of the $2m$ piercing points which have the geometrical property of G_1. In choosing the points of the first row, we may take either of the piercing points on each of the n edges through the corresponding vertex, that is, there are 2^n possible ways to choose the first row. There are but 2^{n-1} ways to choose the second row, for on one of the edges there is but one available piercing point, the other having been already used. Proceeding in this fashion, we see that there are in all

$$2^n \times 2^{n-1} \times 2^{n-2} \times \cdots \times 2^2 \times 2 = 2^m$$

possible groupings of the $2m$ piercing points P_{ij} which have the same geometrical character that G_1 has.

In keeping account of the individual members of this set of 2^m effective groupings it is convenient to use a multiple index system; the general grouping of the set will be denoted by

$$G_\alpha = G_{\alpha_1\alpha_2\alpha_3\cdots\alpha_{n-1}},$$

where the general subscript α_r ranges through $1, 2, 3, \cdots$, to 2^{r+1} for all values of r from 2 to $n-1$ inclusive. The subscript α_1 ranges from 1 to 8. Each of the subscripts controls a certain sub-set of the points P_{ij}; the nature of this control is indicated in the following display:

	α_1	α_2	α_3	α_{n-1}
1	P_{10} P_{20} P_{21}	P_{30} P_{31} P_{32}	P_{40} P_{41} P_{42} P_{43}	P_{n0} $P_{n1}\cdots P_{n,n-1}$
2	P_{10} P_{02} P_{21}	P_{03} P_{31} P_{32}	P_{04} P_{41} P_{42} P_{43}	P_{0n} $P_{n1}\cdots P_{n,n-1}$
3	P_{10} P_{20} P_{12}	P_{30} P_{13} P_{32}	P_{40} P_{14} P_{42} P_{43}	P_{n0} $P_{1n}\cdots P_{n,n-1}$
4	P_{10} P_{02} P_{12}	P_{30} P_{31} P_{23}	P_{40} P_{41} P_{24} P_{43}	⋮
5	P_{01} P_{20} P_{21}	P_{03} P_{13} P_{32}	P_{40} P_{41} P_{42} P_{34}	⋮
6	P_{01} P_{02} P_{21}	P_{03} P_{31} P_{23}	P_{04} P_{14} P_{42} P_{43}	⋮
7	P_{01} P_{20} P_{12}	P_{30} P_{13} P_{23}	P_{04} P_{41} P_{24} P_{43}	⋮
8	P_{01} P_{02} P_{12}	P_{03} P_{13} P_{23}	P_{04} P_{41} P_{42} P_{34}	⋮
⋮			⋮	⋮
16			P_{04} P_{14} P_{24} P_{34}	⋮
⋮				⋮
2^n				P_{0n} $P_{1n}\cdots P_{n-1,n}$.

To get the lower half of the matrix which is the complete display of $G_{\alpha_1\alpha_2\cdots\alpha_{n-1}}$, we adjoin the sets of points P_{ij} controlled by the various indices, and fill in the upper half of the matrix by inversive symmetry. As an example we give the upper left hand corner of the matrix $G_{38}\ldots$:

$$\begin{array}{lcccc} & * & P_{01} & P_{02} & P_{30} \\ \alpha_1 = 3 \rightarrow & P_{10} & * & P_{21} & P_{31} \\ & P_{20} & P_{12} & * & P_{32} \\ \alpha_2 = 8 \rightarrow & P_{03} & P_{13} & P_{23} & * \end{array}.$$

Each of these groupings G_α will define, in the same manner that G_1 did, a simplex, T_α, auxiliary to the fundamental simplex F. The 2^m simplexes T_α are called the auxiliary simplexes of the extended Pascal configuration in S_n, or simply the simplexes of the Kf_n. The 2^m pairs of simplexes composed of the fundamental simplex F and one of the auxiliary simplexes will be called the pairs of Kf_n. It will be shown that the pairs of Kf_n are Schläfli pairs of simplexes, and hence each determines a hyperregulus; the 2^m hyperreguli will be called the hyperreguli of Kf_n; thus the symbol Kf_n means the total configuration of geometric elements associated with the extended Pascal theorem in S_n.

II. Theorems in S_n

Theorem 1. *The intersections of corresponding faces (and the lines joining corresponding vertices) of the pairings of an extended Pascal configuration in S_n are linearly dependent.*

Consider first the pair of simplexes F and T_1, corresponding to the grouping G_1. By the use of (4), it is seen that the equations of the faces of T_1 are

$$\sum_j b_{ij}x_j = 0 \qquad (i,j = 0,1,\cdots,n). \tag{6}$$

The equations of the $n+1$ S_{n-2}'s of intersection with the corresponding faces of F may be written as

$$x_i = \sum_j b_{ij}x_j = 0. \tag{7}$$

The Plücker-Grassmann coördinates of these S_{n-2}'s are the two-rowed determinants formed from the matrices of coefficients in the equations (7). Each set of coördinates has exactly n elements that are not zero; they are shown in the following display:

$$
\begin{array}{lccccccccccc}
 & p_{01} & p_{02} & p_{03}\cdots p_{0n} & p_{12} & p_{13}\cdots p_{1n} & p_{23}\cdots p_{2n}\cdots & p_{n-1,n} \\
\pi_0\rho_0 & b_{01} & b_{02} & b_{03}\cdots b_{0n} & & & & \\
\pi_1\rho_1 & -b_{10} & & & b_{12} & b_{13}\cdots b_{1n} & & \\
\pi_2\rho_2 & & -b_{20} & & -b_{21} & & b_{23}\cdots b_{2n} & \\
\pi_3\rho_3 & & & -b_{30} & & -b_{31} & -b_{32} \quad \cdots & \\
\cdot & & \cdots & \cdots & & & \cdots & b_{n-1,n} \\
\pi_n\rho_n & & & -b_{n0} & & -b_{n1} & -b_{n2}\cdots & -b_{n,n-1}
\end{array} \tag{8}
$$

In each column there are two and only two elements, the sum of which is zero since $b_{ij}=b_{ji}$. The display shows that there is linear dependence between the sets of coördinates of the $n+1$ S_{n-2}'s of intersection of corresponding faces of the simplexes F and T_1, which is what is meant when it is said that the S_{n-2}'s themselves are linearly dependent.

By duality it follows that the lines joining corresponding vertices of the two simplexes are also linearly dependent.

Suppose now the members of certain pairs of piercing points are interchanged. The effect will be to replace the standard grouping G_1 by some particular one of the set G_α; if we know which pairs of piercing points are interchanged, we know which G_α is represented by the modified G_1. For example, suppose that P_{04} and P_{40} are interchanged; the standard grouping G_1 is replaced by $G_{112111\ldots1}$. If now we interchange b_{04} and c_{04} in the equations (7) and display (8), we have converted the proof that G_1 is an effective grouping in Theorem 1 into a proof that the modified grouping is also effective. Since all of the 2^m groupings G_α were obtained from the standard G_1 by such inversions of subscripts among the points P_{ij}, it follows that all the groupings G_α are effective in the theorem.

THEOREM 2. *The pairs of the extended Pascal configuration in S_n are pairs of Schläfli simplexes.*

This follows at once from the theorem* that a set of $n+1$ linearly dependent S_{n-2}'s in S_n are in the position of Schläfli. A quite different proof is based on the fact that the simplexes of a pair are polar reciprocal with respect to a hyperquadric, which, it will be remembered, was Schläfli's original point of departure. The form of equations (6) show that the hyperplanes π_i are the polar hyperplanes of the points P_i with respect to the hyperquadric of equation

$$\sum b_{ij}x_i x_j=0, \tag{9}$$

* Rupp, *A geometrical interpretation of linear dependence in S_n*, abstracted in the Bulletin of the American Mathematical Society, vol. 35 (1929), p. 171.

and accordingly that the points R_i are the poles of the faces ρ_i, that is, the simplexes F and T_1 are polar reciprocal with respect to the hyperquadric in question.

It is a fundamental property of Schläfli simplexes that all the lines which meet n of the S_{n-2}'s of intersection of corresponding faces meet also the remaining S_{n-2}. In the author's note on the V_{n-1}^{n-1} previously cited is given a method of obtaining the point equation of the variety made up of the lines meeting n given S_{n-2}'s of S_n. The use of that method here will give the equations of the varieties bearing the 2^n hyperreguli of the extended Pascal configuration.

By the use of the incidence criterion for line and S_{n-2}, as expressed by Grassmann-Plücker coördinates, the condition that x be the coördinates of a point such that the line joining x to a point of the S_{n-2} $\pi_0\rho_0$ meets also the $n-1$ S_{n-2}'s $\pi_i\rho_i$, where i runs from 1 to $n-1$, is found to be the equation

$$(10)\qquad \begin{vmatrix} b_{01} & b_{02} & b_{03} & \cdots & b_{0,n-1} & b_{0n} \\ -\sum_j^{j\neq 1} b_{1j}x_j & b_{12}x_1 & b_{13}x_1 & \cdots & b_{1,n-1}x_1 & b_{1n}x_1 \\ b_{21}x_2 & -\sum_j^{j\neq 2} b_{2j}x_j & b_{23}x_3 & \cdots & b_{2,n-1}x_2 & b_{2n}x_2 \\ b_{31}x_3 & b_{32}x_3 & -\sum_j^{j\neq 3} b_{3j}x_j & \cdots & b_{3,n-1}x_3 & b_{3n}x_3 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ b_{n-1,1}x_{n-1} & b_{n-1,2}x_{n-1} & \cdot & \cdots & -\sum_j^{j\neq n-1} b_{n-1,j}x_j & b_{n-1,n}x_{n-1} \end{vmatrix} = 0.$$

It should be observed that two options were exercised in writing the foregoing equation, one in choosing the variable line through x to pass through a point of $\pi_0\rho_0$, and one in considering $\pi_n\rho_n$ as the last S_{n-2} of the Schläfli set. The consequence of the first choice is the absence of a variable in the top row of the determinant, and the consequence of the second choice is that the coördinate x_n appears only in the summations. A different ordering of the Schläfli set gives rise to a superficially different form of equation (10), but it is easy to prove the two forms are equivalent.

If we replace, in equation (10), certain of the symbols b_{ij} by the corresponding c_{ij}, the resulting equation will be that of the V_{n-1}^{n-1} bearing the hyperregulus associated with one of the groupings G_α.

THEOREM 3. *The section of the extended Pascal configuration associated with a given hyperquadric Q and fundamental simplex F by a space of the simplex F is itself an extended Pascal configuration plus certain residual flat spaces.*

Consider first the nature of the intersection of a hyperplane of F, say ρ_n, with the V_{n-1}^{n-1} given by equation (10); we shall call the V_{n-1}^{n-1}'s of the Kf_n simply the V_α, in which case the one given by equation (10) is V_1, for it corresponds to G_1. The equation of ρ_n is $x_n = 0$; part of its intersection with V_1 is the $S_{n-2}\,\rho_n\pi_n$, as we may see by a manipulation of the equation of V_1. If the top row of the determinant in (10) is multiplied by x_0, and the other rows added to it, the new top row has the form

$$b_{1n}x_n \quad b_{2n}x_n \quad b_{3n}x_n \quad \cdots \quad b_{n-1,n}x_n \quad b_{0n}x_0 + b_{1n}x_1 + \cdots + b_{n-1,n}x_{n-1},$$

whose elements are identically zero if x satisfies the restrictions

$$x_n = \sum_i b_{in}x_i = 0,$$

that is, if x lies on $\rho_n\pi_n$. We remark that this incidentally furnishes an analytic proof of the theorem of Wong that the S_{n-2}'s defining a V_{n-1}^{n-1} lie on it; in this connection it may be said that the same theorem follows at once from the present writer's geometric interpretation of linear dependence in S_n, for, just as $n+2$ points which are linearly dependent lie in, and determine, a flat space, an S_n, so do $n+1$ linearly dependent S_{n-2}'s lie in, and determine, a curved manifold, a V_{n-1}^{n-1}.

To return to the discussion of Theorem 3, we have seen that one factor of the intersection of V_1 and ρ_n is an S_{n-2}; it remains to find the residual portion. Suppose a point of ρ_n is such that its coördinates annul the determinant which is the cofactor of $b_{n-1,\,n}x_{n-1}$ in the left member of (10); it will lie on a V_{n-2}^{n-2}, for it satisfies an equation of the same form as (10). We do not mean to say that the V_{n-2}^{n-2} and the S_{n-2} of intersection of V_1 with a hyperplane of F have no common points; they actually meet in a V_{n-3}^{n-2}. For our present purposes, nevertheless, it is the V_{n-2}^{n-2} and the S_{n-2} which are the important spaces of intersection.

We have seen that an S_{n-1} of F meets V_1 in a V_{n-2}^{n-2}. Consider now the intersection of the 2^n V_{n-1}^{n-1}'s whose symbols are

$$V_{11\cdots 11\alpha_{n-1}} \qquad (\alpha_{n-1} = 1, 2, \cdots, 2^n),$$

with the $S_{n-1}\,\rho_n$. They all share the V_{n-2}^{n-2} whose equation, in ρ_n, is obtained by setting the cofactor of $b_{n-1,\,n}x_{n-1}$ in the determinant in the left member of (10) equal to zero, because the index α_{n-1} by definition controls only the symbols b_{ni}, where i runs from 0, 1, to $n-1$, which appear exclusively, if we keep to the $S_{n-1}\,\rho_n$, in the last column of the determinant of (10). Varying the first $n-2$ indices of V_α is equivalent to replacing certain of the b_{ij} by the corresponding c_{ij}. Thus ρ_n meets the 2^n V_{n-1}^{n-1}'s which bear the Kf_n in 2^{n+1} V_{n-2}^{n-2}'s, and these V_{n-2}^{n-2}'s bear the Kf_{n-1} of the simplex $P_0P_1\cdots P_{n-1}$

set up by means of the hyperquadric which is the intersection of Q with ρ_n. The $S_{n-1}\,\rho_n$ also meets the V_{n-1}^{n-1}'s of the Kf_n in the S_{n-2}'s in which it meets its corresponding faces among the auxiliary simplexes $T_{11\cdots 11\alpha_{n-1}}$. We may clearly reduce the Kf_{n-1} to a Kf_{n-2} in a similar manner. Had we desired to ascertain the nature of the intersections of the Kf_n with another S_{n-1} of F, we could clearly have chosen the appropriate form of equation (10) and proceeded as above.

An example of the successive reduction of the Kf_n is given below where we show that the planes of a tetrahedron meet the 64 V_2^2's bearing the Kf_3 in a Kf_2 and certain residual spaces.

III. The extended Pascal configuration in the plane

For the case that n is 2, we consider a fundamental triangle F and conic Q. Paired with F are the triangles

$$T_{\alpha_1} \qquad (\alpha_1 = 1,2,\cdots,8),$$

each pairing of triangles possessing the property that the points of intersection of corresponding sides are linearly dependent, i.e., collinear. The line of collinearity is an axis of perspectivity for the pair of triangles, and also the ordinary Pascal line of the Pascal hexagon whose sides are the sides of the two triangles, arranged in proper order. The 2^m V_{n-1}^{n-1}'s here reduce to the eight lines V_{α_1}; the line V_{α_1} associated with an effective grouping G_{α_1} constitutes the entire class of lines meeting any two of the three linearly dependent Schläfli points of the pair of triangles, and as a member of such a class it has the Schläfli property of meeting the third point. The Kf_2 consists of these eight V_{α_1}; the reciprocals of the line coördinates of these are shown in the array below:

(11)

	V_1	V_2	V_3	V_4	V_5	V_6	V_7	V_8
$1/u_1$	b_{12}	b_{12}	c_{12}	c_{12}	b_{12}	b_{12}	c_{12}	c_{12}
$1/u_2$	b_{20}	c_{20}	b_{20}	c_{20}	b_{20}	c_{20}	b_{20}	c_{20}
$1/u_3$	b_{01}	b_{01}	b_{01}	b_{01}	c_{01}	c_{01}	c_{01}	c_{01}.

We notice that the pairs of lines whose indices sum to nine intersect upon the line of equation

$$\sum_k \left(\frac{1}{b_{ij}} + \frac{1}{c_{ij}}\right) x_k = 0 \qquad (i,j,k \text{ a permutation of } 0,1,2), \tag{12}$$

an equation which may be rewritten, by the use of (3), in the form

$$\sum_k \frac{a_{ij}}{a_{ii}a_{jj}} x_k = 0. \tag{12'}$$

This line is a Steiner-Plücker line of the Pascal hexagon whose vertices are the six points P_{ij}; the three hexagons formed of the triangle pairs FT_1, T_1T_8, and T_8F, are readily seen to be three hexagons in the Steinerian relation, i.e., three vertices are fixed, and the remaining three permuted cyclically. It follows that the lines V_1 and V_8 meet in a Steiner point; since the same argument applies to all pairs of the Kf_2 whose indices sum to nine, the four points which define the line of equation (12) must be Steiner points, and hence their line of collinearity is a Steiner-Plücker line.

An independent check is found by using a table due to Cayley* which shows the nature of the intersection of the pairs of the 60 Pascal lines. In correlating the two notations, let the triangles F and T_1 be the hexagon denoted by the Cayley letters AE. The sides are arranged in the order $\rho_2\pi_0\rho_1\pi_2\rho_0\pi_1$, and the vertices in the order $P_{01}P_{02}P_{20}P_{21}P_{12}P_{10}$, or, for brevity, in the order 123456. If the symbols p, h, and g denote Pascal, Kirkman, and Steiner points respectively, Cayley's table can be correlated with the present paper as below:

Group	Hexagon	Cayley Letters		Nature of Intersection						
				V_1	V_2	V_3	V_4	V_5	V_6	V_7
1	123456	AE								
2	132456	EL	V_2	p						
3	123546	EF	V_3	p	h					
4	132546	EG	V_4	h	p	p				
5	162345	DE	V_5	p	h	h	g			
6	163245	EI	V_6	h	p	g	h	p		
7	162354	EM	V_7	h	g	p	h	p	h	
8	163254	EH	V_8	g	h	h	p	h	p	p

This display shows that the pairs V_1V_8, V_2V_7, V_3V_6, and V_4V_5 meet in Steiner points, but it does not show that the Steiner points thus garnered lie on a line.

THEOREM 4. *The Kf_2 determines a Steiner-Plücker line.*

In the plane, the Kf_2 is apparently a less rich configuration than the Hexagrammaticum Mysticum, for it has but 8 lines where the other has 60. Consider, however, a different approach to the general situation in S_n by taking, as did Schläfli, two simplexes polar reciprocal with respect to a hyperquadric. We have already observed that F and T_1 are polar reciprocal

* Cayley, *A notation for Pascal's theorem*, Quarterly Journal of Mathematics, vol. 9 (1868), pp. 268–274.

with respect to the hyperquadric given by equation (9). Let us take this hyperquadric as basic, and try to determine the hyperquadric Q whose matrix involves the numbers a_{ij}. We now define the set of 2^m points P_{ij} as the intersection of P_iP_j with π_i, and an analogous set of 2^m points R_{ij} as the intersection of R_iR_j with ρ_i; it is an easily proved property of Schläfli simplexes that the points P_{ij} and R_{ij} lie upon hyperquadrics, which are in general different, coinciding only when n is 2. In the plane there is associated with a fundamental triangle F by means of a conic Q a set of eight auxiliary triangles; associated with any one of these auxiliary triangles by means of the same conic Q is a set of eight more auxiliary triangles, which set partially overlaps the first. Indeed, the six points P_{ij} where the sides of F meet Q determine 15 lines, which may be grouped, by triples, into 15 distinct triangles. Any one of these 15 triangles pairs with eight others to make a Pascal hexagon; there are $(15\times 8)/(1\times 2)=60$ different Pascal hexagons. It is because the points R_{ij} coincide with the points P_{ij} in the plane case that there exists a configuration of more than 2^n lines. The richness of the usual Pascal configuration is from this point of view due to the coincidence of the hyperquadrics (in the plane case, conics) determined by the sets of points P_{ij} and R_{ij}.

The body of theorems about the Pascal configuration can be established from a consideration of the 15 triangles instead of the more usual approach by way of the 60 Pascal hexagons.

IV. The extended Pascal configuration in space

The edges of a fundamental tetrahedron F pierce a given quadric (1) in the 12 points P_{ij}, which are grouped in the 64 effective groupings $G_{\alpha_1\alpha_2}$ $(\alpha_1\alpha_2=1, 2, \cdots, 8)$, each of which groupings defines an auxiliary tetrahedron $T_{\alpha_1\alpha_2}$. The lines of intersection of corresponding faces of the fundamental tetrahedron F and an auxiliary tetrahedron $T_{\alpha_1\alpha_2}$ are linearly dependent, hence these four lines lie on a regulus. The Kf_3 is made up of the 64 conjugate reguli, that is, of the 64 one-parameter families of lines which meet the 64 sets of linearly dependent lines. A pairing of the Kf_3, as F and $T_{\alpha_1\alpha_2}$, is a Schläfli pair of tetrahedra, that is, such that all the lines meeting three of the lines of intersection of corresponding faces meet also the fourth.

The theorem of Chasles and Weddle asserts that the four lines of intersection of corresponding faces lie on a regulus. Chasles further remarked that there are several effective groupings of the 12 piercing points, but apparently he did not consider how many effective groupings there are, nor what incidence relations exist among the various reguli.

The V_{n-1}^{n-1} defined by a pairing is here a V_2^2, or quadric; it contains the regulus of lines which meet the four linearly dependent lines, and also the lines themselves.

THEOREM 5. *The fundamental tetrahedron F circumscribes each of the quadrics* $V_{\alpha_1\alpha_2}$.

In the derivation of $V_{\alpha_1\alpha_2}$ it appears that it contains the lines $\rho_i\pi_i$ $(i=0, 1, 2, 3)$, that is, it has a generator in each of the planes ρ_i, hence it is tangent to each of these planes.

THEOREM 6. *The* 128 *lines common to a face of F and the* 64 *quadrics* $V_{\alpha_1\alpha_2}$ *consist of two* Kf_2*'s, each line being counted eight times.*

The face ρ_3 has the equation $x_3=0$. It meets V_{11} in a conic of equation

$$0 = x_3 = \Big(\sum_i^{012} b_{ij}b_{ik}x_i \Big)\Big(\sum_i^{012} b_{i3}x_i \Big), \tag{13}$$

where i, j, k is a permutation of 0, 1, 2. The eight quadrics $V_{1\alpha_2}$ have equations differing from that of V_{11} only in the coefficients controlled by α_2, i.e., the coefficients b_{03}, b_{13}, and b_{23}, which appear segregated in one factor of (13). It follows that the eight quadrics $V_{1\alpha_2}$ share the ruling of equations

$$0 = x_3 = \sum_i^{012} b_{ij}b_{ik}x_i. \tag{14}$$

In like manner the eight quadrics $V_{2\alpha_2}$ share the ruling of equations

$$0 = x_3 = b_{01}c_{02}x_0 + b_{10}b_{12}x_1 + c_{20}b_{21}x_2.$$

The eight lines obtained by varying the first index constitute the Kf_2 of $P_0P_1P_2$ with respect to the conic cut from Q by $x_3=0$, as may be seen by comparing their equations with the display (11).

Octuples of quadrics sharing a common second index likewise have a common generator. The coördinates, in $x_3=0$, of the eight lines thus obtained are shown below:

$$\begin{array}{cccccccccc} & L_1 & L_2 & L_3 & L_4 & L_5 & L_6 & L_7 & L_8 \\ u_1 & b_{03} & c_{03} & b_{03} & b_{03} & c_{03} & c_{03} & b_{03} & c_{03} \\ u_2 & b_{13} & b_{13} & c_{13} & b_{13} & c_{13} & b_{13} & c_{13} & c_{13} \\ u_3 & b_{23} & b_{23} & b_{23} & c_{23} & b_{23} & c_{23} & c_{23} & c_{23} \end{array} \tag{15}$$

The general similarity of (11) and (15) suggests that the second set of eight

lines is the Kf_2 of $P_0P_1P_2$ with respect to some conic. To find this conic, solve (3) for a_{ij} in terms of b_{ij}, obtaining

$$a_{ii} = b_{ii}, \quad 2b_{ij}a_{ij} = b_{ij}^2 + b_{ii}b_{jj}.$$

We next consider how we might pass from the lines of (11) to the conic whose matrix involves the numbers a_{ij}, and use the same method to obtain the conic of matrix

$$(16) \qquad \begin{pmatrix} 2a_{00} & \beta_{23} + a_{00}a_{11}b_{23} & \beta_{13} + a_{00}a_{22}b_{13} \\ \beta_{23} + a_{00}a_{11}b_{23} & 2a_{11} & \beta_{03} + a_{11}a_{22}b_{23} \\ \beta_{13} + a_{00}a_{22}b_{13} & \beta_{03} + a_{11}a_{22}b_{03} & 2a_{22} \end{pmatrix},$$

where $\beta_{ij}b_{ij} = 1$, by means of which the lines of coördinates (15) are the Kf_2 of $P_0P_1P_2$. This conic appears to have no simple geometrical relation to the original quadric Q, whereas the corresponding conic associated with the first set of eight lines is the intersection of Q by $x_3 = 0$.

It is clear that ρ_3 is in no wise an exceptional face of F, hence the situation on it is duplicated on the other planes of the tetrahedron. This completes the proof of Theorem 6.

THEOREM 7. *A tetrahedron and a quadric determine two sets of Steiner-Plücker lines, each set lying on a regulus.*

On each face of the fundamental tetrahedron there are two Steiner-Plücker lines, one deriving from each of the Kf_2's. One Kf_2 consists of lines belonging to the reguli which form the Kf_3 of the tetrahedron and quadric; the four associated Steiner-Plücker lines have the equations

$$(17) \qquad 0 = x_i = \sum_j^{j \neq i} \frac{a_{km}}{a_{kk}a_{mm}} x_j \qquad (ijkm \text{ a permutation of } 0123),$$

equations which show by their symmetry that the four lines in question lie on a regulus. The same is true of the equations of the four Steiner-Plücker lines deriving from the Kf_2's whose lines lie on the conjugate reguli of the Kf_3; the equations are

$$(17') \qquad 0 = x_i = \sum_j^{j \neq i} (b_{ji} + c_{ji})x_j.$$

V. THE EXTENDED PASCAL CONFIGURATION IN S_4

Here we have 1024 two-parameter families of lines meeting the quintuples of planes of intersection of corresponding hyperplanes in the pairings of a

fundamental simplex F with the auxiliary simplexes $T_{\alpha_1\alpha_2\alpha_3}$ ($\alpha_1, \alpha_2 = 1, 2, \cdots, 8$, $\alpha_3 = 1, \cdots, 16$), these auxiliary simplexes being defined by the 1024 effective groupings $G_{\alpha_1\alpha_2\alpha_3}$ of the 20 points in which the edges of F pierce a given hyperquadric Q. The two-parameter families of lines, or hyperreguli, lie on a set of V_3^3's; the V_3^3 is a V_{n-1}^{n-1} which has been thoroughly studied by Segre* and Castelnuovo,† its co-discoverers, Berzolari,‡ and many others.

The 16 V_3^3's denoted by $V_{11\alpha_3}$ meet the S_3 whose equation is $x_4 = 0$ in the same quadric; its equation in ρ_4 is that of V_{11} in three-space. It follows that the 1024 V_3^3's meet ρ_4 in $64 V_2^2$'s, each counted 16 times; this is a special case of Theorem 3.

---

* Segre, Atti, Accademia delle Scienze di Torino, vol. 22 (1887), pp. 791–801; Memorie, Accademia delle Scienze di Torino, (2), vol. 39 (1888), pp. 3–32.

† Castelnuovo, Atti, Istituto Veneto, (6), vol. 5 (1887), p. 1249, and vol. 6 (1888), p. 525.

‡ Berzolari, Rendiconti, Accademia dei Lincei, (5), vol. 25_1 (1917), pp. 29 and 102.

PENNSYLVANIA STATE COLLEGE,
STATE COLLEGE, PA.

CONCERNING THE ARC-CURVES AND BASIC SETS OF A CONTINUOUS CURVE, SECOND PAPER*

BY

W. LEAKE AYRES†

I. INTRODUCTION

In an earlier paper‡ with the same title, we have defined and studied the properties of certain subsets of a continuous curve§ which we call the *arc-curves* of the continuous curve. In a recent paper, G. T. Whyburn‖ has defined the *cyclic elements* of a continuous curve, and he has considered a continuous curve as composed of its cyclic elements and has given a large number of the properties of connected collections of cyclic elements. On examining the two papers it is found that arc-curves and connected collections of cyclic elements have many properties in common; and, in fact, in part II of the present paper we shall show that, although these two sets were defined very differently, every connected collection of cyclic elements of a continuous curve is an arc-curve of the continuous curve, and conversely, every arc-curve that contains more than one point is a collection of cyclic elements of the continuous curve. In part III we will develop some new theory concerning the basic sets of a continuous curve, which were defined in *Arc-curves, first paper*, and shall show the relation between the basic sets and the nodes of a continuous curve. In part IV we shall show that an irreducible basic set of a continuous curve resembles in its properties the set of all end points of the continuous curve.

All point sets considered in this paper are assumed to lie in a metric, separable, locally compact space.

Notation. We shall use the common notation of the theory of sets, such as $A+B$, $A-B$, $A \cdot B$, etc., in its usual meaning. If H is a point set, the symbol $\overline{H}$ denotes the point set consisting of the points of H together

---

* Presented to the Society, San Francisco Section, June 2, 1928; received by the editors in June, 1928.

† National Research Fellow in Mathematics.

‡ *Concerning the arc-curves and basic sets of a continuous curve*, these Transactions, vol. 30 (1928), pp. 567–578. Hereafter in the present paper we will refer to this paper as *Arc-curves, first paper*.

§ A *continuous curve* is a point set which is closed, connected and connected im kleinen. Unless definitely stated, there is no assumption as to compactness.

‖ *Concerning the structure of a continuous curve*, American Journal of Mathematics, vol. 50 (1928), pp. 167–194. Hereafter we will refer to this paper as *Structure*.

with all the limit points of H. If X and Y are point sets (either of which may, in particular, be a single point), the *distance* from X to Y, denoted by $d(X, Y)$, is the greatest lower bound of all the numbers $d(x, y)$, where x is a point of X, y is a point of Y and $d(x, y)$ denotes the distance from the point x to the point y. If xyz denotes an arc with end points x and z, the symbols $<xyz, xyz>, <xyz>$ denote $xyz-x$, $xyz-z$ and $xyz-x-z$ respectively.

II. The arc-curves and the cyclic elements

Definitions.* If K is a subset of the continuous curve M, the set of all points $[P]$, such that P lies on some arc of M whose end points belong to K, is called the *arc-curve of K with respect to M* and is denoted by $M(K)$. In case the set K consists of but a single point, the arc-curve of K with respect to M is defined to be this single point. A continuous curve M is said to be *cyclicly connected* if every two points of M lie together on some simple closed curve of M. A cyclicly connected continuous curve C which is a subset of a continuous curve M is said to be a *maximal cyclic curve* of M if and only if no cyclicly connected continuous curve which is a subset of M contains C as a proper subset. A subset E of a continuous curve M is said to be a *cyclic element* of M provided that E is either (a) a maximal cyclic curve of M, (b) a cut point† of M or (c) an end point‡ of M. A point set N is said to be a *simple cyclic chain* of M between two cyclic elements E_1 and E_2 of M provided that N is connected, contains E_1 and E_2, is the sum of the elements of some collection of cyclic elements of M, and furthermore no proper connected subset of N contains E_1 and E_2 and is the sum of the elements of such a collection.

Theorem 1. *If M is a continuous curve and K is a subset of M containing more than one point, then the arc-curve $M(K)$ is a collection of cyclic elements of M. Conversely, if the connected subset H of M is a collection of cyclic elements of M, then there is a subset K of H such that H is the arc-curve $M(K)$.*

---

* These definitions are given in *Arc-curves, first paper*, p. 568, and *Structure*, pp. 167–8.

† The point P of the connected set M is said to be a *non-cut point* or a *cut point* according as the set $M-P$ is connected or not. See R. L. Moore, *Concerning the cut points of continuous curves and other closed and connected point sets*, Proceedings of the National Academy of Sciences, vol. 9 (1923), pp. 101–106.

‡ The point P of a continuous curve M is said to be an *end point* of M provided that if P' is any other point of M and $P'P$ is any arc of M with end points P and P', then the set $M-P'P>$ contains no connected subset containing more than one point that contains the point P. See R. L. Wilder, *Concerning continuous curves*, Fundamenta Mathematicae, vol. 7 (1925), p. 358. For other definitions that are equivalent to that of Wilder, see H. M. Gehman, *Concerning end points of continuous curves and other continua*, these Transactions, vol. 30 (1928), pp. 63–84.

Let A and B be any two points of K and let α be any arc of M from A to B. Let P be any point of the set $<\alpha>$. If P is a cut point of M, then it is itself a cyclic element of M. If P is not a cut point of M, then by a result due to G. T. Whyburn* there is a maximal cyclic curve C having in common with α an arc β which contains P. Let X and Y denote the end points of β such that we have the order $AXYB$ on the arc α. If Q is any point of $C-X-Y$, it follows by a theorem due to the author† that there exists an arc γ of C with end points X and Y and containing the point Q. Then the arc γ together with the subarcs AX and YB of α is an arc of M with end points A and B. Hence every point Q of C belongs to the arc-curve $M(K)$. Thus if P is not a cut point of M, it belongs to a maximal cyclic curve of M which belongs to the arc-curve $M(K)$.

Now consider the point A, the case for B being similar. If A is an end point or a cut point of M, then it is itself a cyclic element of M. If the point A is neither, it belongs to some simple closed curve J which is a subset of M.‡ There is a maximal cyclic curve C containing the simple closed curve J.§ As A is a non-cut point of M, the curve C and the arc α have in common an arc β containing A.‖ Let X be the end point of β different from A. If Q is any point of $C-A$, there exists an arc γ of C with end points A and X and containing the point Q. Then the arc γ plus the subarc XB of α is an arc of M whose end points A and B belong to K. Hence every point Q of C belongs to the arc-curve $M(K)$.

We have shown that every point of the arc-curve $M(K)$ is a subset of

---

* G. T. Whyburn, *Some properties of continuous curves*, Bulletin of the American Mathematical Society, vol. 33 (1927), pp. 305–308. Whyburn states the result for two dimensions only, but with the use of certain results established in my paper *Concerning continuous curves in metric space*, to appear in the American Journal of Mathematics, it may be easily extended to the more general space considered here.

† See my paper *Continuous curves which are cyclicly connected*, Bulletin de l'Académie Polonaise des Sciences et des Lettres, 1928, pp. 127–142.

‡ See W. L. Ayres, *Concerning continuous curves and correspondences*, Annals of Mathematics, vol. 28 (1927), pp. 396–418, Theorem 3, and G. T. Whyburn, *Concerning continua in the plane*, these Transactions, vol. 29 (1927), pp. 369–400, Theorem 22. These results are extended to more general space in my paper *Concerning continuous curves in metric space*. In general this is true of all other references in this paper which state results for two dimensions only. When this is not the case mention will be made of the extension. Otherwise it will be understood that the extension is to be found in the above paper.

§ G. T. Whyburn, *Cyclicly connected continuous curves*, Proceedings of the National Academy of Sciences, vol. 13 (1927), pp. 31–38, Theorem 3.

‖ If P is a non-cut point of a continuous curve M belonging to a maximal cyclic curve C of M, and α is an arc of M containing P, then C and α have in common an arc β containing P, and if P is interior to the arc α it is an interior point of β.

The proof is sufficiently evident to be omitted.

some cyclic element of M that belongs to the arc-curve $M(K)$; or in other words, the arc-curve $M(K)$ is a collection of cyclic elements of M.

Conversely, we shall show that if H is a connected subset of M which is a collection of cyclic elements of M, then H contains a subset K such that H is the arc-curve $M(K)$. Let K be the set H. G. T. Whyburn* has shown that H contains every arc of M joining two points of H; hence the arc-curve $M(H)$ is a subset of H. But evidently H is a subset of the arc-curve $M(H)$. Therefore H is identical with the arc-curve $M(H)$. There are also many proper subsets K of H such that H is identical with the arc-curve $M(K)$.

THEOREM 2. *If A and B are distinct points of the continuous curve M, and K is the arc-curve $M(A+B)$, then there are two cyclic elements E_1 and E_2† of M containing A and B respectively such that K is a simple cyclic chain of M between the two cyclic elements E_1 and E_2. Conversely, if H is a simple cyclic chain of M between two cyclic elements E_1 and E_2 of M, then E_1 and E_2 contain points A and B respectively such that H is the arc-curve $M(A+B)$.*

By Theorem 1 there are two cyclic elements E_1 and E_2 of M that contain A and B respectively and belong to the arc-curve K. Also the set K is connected and is the sum of the elements of a collection of cyclic elements of M. There is a unique‡ simple cyclic chain H of M between E_1 and E_2 which is a continuous curve§ and is a subset of K.‖ If H is a proper subset of the continuous curve K,¶ let N be a component of $K-H$. If the component N of $K-H$ has more than one limit point in H, it is easy to see that M contains an arc whose end points belong to H and which lies except for its end points entirely in the set N. But this is impossible since H contains every arc of M joining two points of H.** Hence the set H contains just one limit point of N. Let P be this point. Then P is a cut point of the arc-curve K and the component N of $K-P$ contains neither A nor B. This is impossible by a theorem of the author.†† The assumption that H is a proper subset of K leads to an absurdity. Thus H is identical with K.

For the converse, if E_1 is a single point let this point be A. If E_1 is a

---

* *Structure*, Theorem 8.

† The two elements E_1 and E_2 are not necessarily different.

‡ *Structure*, Theorem 3.

§ *Structure*, Theorem 4.

‖ *Structure*, Theorem 8.

¶ Whenever X is a closed point set, the arc-curve $M(X)$ is a continuous curve. See *Arc-curves, first paper*, Theorem 7.

** *Structure*, Theorem 8.

†† If P is a cut point of the arc-curve $M(K)$, every component of $M(K)-P$ contains a point of K. See *Arc-curves, first paper*, Theorem 9.

maximal cyclic curve of M, let A be any point of E_1 that is not a cut point of H. Only one point of E_1 is a cut point of H and thus A may be chosen as any other point of E_1. In a similar manner let us select a point B of the cyclic element E_2. Since every arc of M joining two points of H belongs to H, the arc-curve $M(A+B)$ is a subset of H; and in much the same manner as above we may show that it is not a proper subset. Then the arc-curve of $A+B$ with respect to M must be identical with the simple cyclic chain H.

In view of the relation between collections of cyclic elements and arc-curves shown in the preceding theorems, we have the following forms for three results of G. T. Whyburn.*

THEOREM 3. *If A and B are distinct points of the continuous curve M, K is the arc-curve $M(A+B)$ and G is a cyclic element of K that contains neither A nor B, then $K-G=K_1+K_2$, where K_1 and K_2 are mutually separated point sets, and* (1) *K_1 and K_2 are connected and contain A and B respectively,* (2) *the cyclic element G contains a point C such that K_1+G and K_2+G are the arc-curves of $A+C$ and $B+C$ respectively with respect to M.*

THEOREM 4. *If A and B are distinct points of the continuous curve M, the arc-curve H of $A+B$ with respect to M is not disconnected by the omission of any cyclic element of M which contains either A or B and belongs to H, or by any set of such cyclic elements.*

THEOREM 5. *Every arc-curve of the continuous curve M is uniformly connected im kleinen.*

Since every connected collection of cyclic elements is an arc-curve, and every arc-curve of two points is a simple cyclic chain, and the converse of each is true, it is to be expected that there is some overlap in the results of the two papers. For instance, Whyburn† proves that if H is a connected collection of cyclic elements of M, then (1) H is arc-wise connected, cyclicly chainwise connected‡ and contains every simple continuous arc of M which joins two points of H, (2) each point of $\overline{H}-H$ is a cyclic element, (3) if H is closed it is a continuous curve. Property (1) is contained in Theorem 3 of *Arc-curves, first paper*, which states that if K is any subset of M and α is an arc of M whose end points belong to $\overline{M(K)}$, then the arc-curve $M(K)$ contains every point of α except possibly the end points. In *Arc-curves*,

* *Structure*, Theorems 5, 6, 9.

† *Structure*, Theorems 8, 10, 11.

‡ A subset H of a continuous curve M is said to be *cyclicly chainwise connected* if every two cyclic elements of M which belong to H can be joined by a simple cyclic chain of M that is a subset of H. See *Structure*, p. 175.

first paper we have the result that every point of $\overline{M(K)} - M(K)$ is an end point of the continuous curve $\overline{M(K)}$. This corresponds closely to property (2). Property (3) is identical with Theorem 6 of *Arc-curves, first paper*.

Menger* has shown that a regular curve B is an acyclic continuous curve if and only if the common part of any two subcontinua of B is connected or vacuous. We may state a similar property of the arc-curves of any continuous curve.

THEOREM 6. *If K and H are two subsets of a continuous curve M, the set of common points of the arc-curves $M(H)$ and $M(K)$ is either vacuous or itself an arc-curve of M.*

Let N be the set of common points of the arc-curves $M(H)$ and $M(K)$. If N is vacuous or contains but a single point, our theorem is true, for a point is the arc-curve of that point with respect to M. Now let x and y be two distinct points of N, and let α be an arc of M with end points x and y. By a theorem of the author,† every point of an arc joining two points of an arc-curve belongs to the arc-curve. Hence every point of α belongs to both $M(K)$ and $M(H)$, and thus N contains the arc α. Thus every arc of M joining two points of N belongs to N. Hence N is its own arc-curve with respect to M.

Expressed in terms of cyclic elements, Theorem 6 is as follows: *The common part of any two connected collections of cyclic elements of the continuous curve M is either vacuous or itself a connected collection of cyclic elements of M.*

III. THE NODES AND BASIC SETS OF A CONTINUOUS CURVE

Definitions. The subset K of the continuous curve M is said to be a *basic set* of M if M is identical with the arc-curve $M(K)$. The subset K of M is said to be an *irreducible basic set* of M if K is a basic set of M but no proper subset of K is a basic set of M.‡ A subset E of a continuous curve M is said to be a *node* of M if E is an end point of M or is a maximal cyclic curve of M that contains at most one cut point of M.§ If H is a connected collection of cyclic elements of the continuous curve M, the subset E of H will be said to be a *node* or an *end point* of H provided it is a node or an end point of the continuous curve $\overline{H}$.

---

* *Ueber reguläre Baumkurven*, Mathematische Annalen, vol. 96 (1926), p. 575.

† *Arc-curves, first paper*, Theorem 3.

‡ For these definitions, see *Arc-curves, first paper*, p. 575.

§ See *Structure*, p. 178.

THEOREM 7.* *In order that a cyclic element E of a continuous curve M should be a node of M it is necessary and sufficient that the set K should contain a point of E which is a non-cut point of M, for any arc-curve $M(K)$ that contains the set E.*

The condition is necessary. Let $M(K)$ be any arc-curve of M containing the set E. If E is an end point of M, no arc of M contains E as an interior point.† Hence, as E belongs to the set $M(K)$, it must be a point of the set K. As every end point of the continuous curve M is a non-cut point of M, our condition is satisfied. If E is a maximal cyclic curve of M containing no cut point of M, then $M=E$ and every point of K is a non-cut point of M and belongs to E. If E is a maximal cyclic curve of M containing just one cut point P of M, then either our condition is satisfied or $E-P$ contains no point of the set K. Suppose the latter holds. Then the set K is a subset of the set $M-(E-P)$. It is evident that

$$M = E + (M - (E - P)),$$

and that the sets E and $(M-(E-P))$ have just one point in common. Then the set $M-(E-P)$ contains every arc of M whose end points belong to it. As K is a subset of $M-(E-P)$, the arc-curve $M(K)$ is a subset of $M-(E-P)$. But this is impossible as the arc-curve $M(K)$ contains the set E.

The condition is sufficient. Suppose the cyclic element E is not a node of M. Then E is either a cut point or a maximal cyclic curve of M containing more than one cut point of M. If E is a cut point of M, let x and y be points of $M-E$ that do not belong to the same component of $M-E$. Then every arc of M with end points x and y contains E. Thus the arc-curve $M(x+y)$ contains E but the set $x+y$ belongs to $M-E$. This is contrary to our assumed condition. If E is a maximal cyclic curve of M containing more than one cut point of M, let x and y be two cut points of M that belong to E. As every point of E lies on some arc of E with end points x and y‡, the arc-curve $M(x+y)$ contains every point of E. But neither x nor y is a non-cut point of M, which contradicts our assumed condition. The assumption that E is not a node has led to a contradiction with our condition in both cases. Hence the condition is sufficient.

---

* Compare this result with Theorem 13 of *Structure*.

† See G. T. Whyburn, *Concerning continua in the plane*, loc. cit., Theorem 12, and my paper *Concerning continuous curves and correspondences*, loc. cit., Theorem 4.

‡ See my paper *Continuous curves which are cyclicly connected*, loc. cit.

THEOREM 8. *In order that a subset K of a compact continuous curve M, which is not cyclicly connected, be a basic set of M it is necessary and sufficient that each node of M should contain at least one non-cut point of M that belongs to K.*

The condition is necessary by the preceding theorem. We shall show that the condition is also sufficient. Let P be any point of M. If P is an end point of M, then P belongs to K, for every end point is a node. If P is a cut point of M, let M_1 and M_2 be two components of $M-P$. If M_1+P is cyclicly connected, then M_1+P is a node of M and M_1 contains a point of K by the condition. If M_1+P is not cyclicly connected, then M_1+P contains two nodes of itself.* There are three cases to consider: (a) one of the nodes is an end point z of M_1+P distinct from P, (b) P is an end point of M_1+P and the other node is a maximal cyclic curve C of M_1+P, (c) both nodes are maximal cyclic curves of M_1+P. Under case (a) if there exists an arc of M containing the point z as an interior point, then this arc contains a subarc containing z as an interior point and belonging to M_1+P, for there is a neighborhood of z such that every point of M belonging to this neighborhood belongs to M_1+P. But, as z is an end point of M_1+P, no arc of M_1+P contains z as an interior point. Thus no arc of M has z as an interior point and z is an end point of M. Hence z belongs to K by our condition. In case (b) let Q be the cut point of M_1+P belonging to the node C. The point P does not belong to C since P is an end point of M_1+P and hence belongs to no simple closed curve of M_1+P. Then

$$M_1 + P - Q = (C - Q) + M_P,$$

where $C-Q$ and M_P are mutually separated sets and M_P contains the point P. As $M-(M_1+P)$ has just one limit point P in M_1+P, the sets $C-Q$ and $M-C$ are mutually separated. Then the set C is a node of M. Then $C-Q$ contains a point of K by our condition. In case (c) let C_1 and C_2 be two nodes of M_1+P which are maximal cyclic curves of M_1+P. If either C_1 or C_2 does not contain P, then, as in case (b), that maximal cyclic curve is a node of M and contains a point of K by the condition. If both C_1 and C_2 contain P, then P is a cut point of M_1+P as every point common to two maximal cyclic curves of a continuous curve is a cut point of that curve. But P is not a cut point of M_1+P, for the set M_1 is connected. Thus in any of the three cases we see that M_1 contains a point x of K. Similarly the component M_2 of $M-P$ contains a point y of K. Every arc of M whose end

* See *Structure*, Theorem 14.

points are x and y contains the point P. Hence every cut point of M belongs to the arc-curve $M(K)$.

If P is neither a cut point nor an end point of M, then P is a non-cut point of M which belongs to some simple closed curve J of M. Let C be the maximal cyclic curve of M containing the curve J. As M is not cyclicly connected the maximal cyclic curve C contains at least one cut point of M. If C contains just one cut point Q of M, then C is a node of M and contains a non-cut point y of M belonging to K. If this point is P, then the arc-curve $M(K)$ contains P. If not, let M_1 be a component of $M-Q$ not containing $C-Q$. As shown in the preceding paragraph, the component M_1 of $M-Q$ contains a point x of K. There exists an arc of C with end points Q and y and containing P, and there exists an arc with end points Q and x and lying wholly in M_1+Q. The sum of these two arcs is an arc of M containing P and whose end points belong to K. Hence P belongs to the arc-curve $M(K)$. If C contains more than one cut point of M, let Q_1 and Q_2 be two such points. Let M_1 and M_2 be components of $M-Q_1$ and $M-Q_2$ respectively that do not contain $C-Q$. Evidently M_1 and M_2 are mutually separated. By the argument of the preceding paragraph we see that M_1 and M_2 contain points x and y of K. There exists an arc of C with end points Q_1 and Q_2 and containing the point P, and there exist two arcs with end points x and Q_1 and y and Q_2 and lying in M_1+Q_1 and M_2+Q_2 respectively. The sum of the three arcs is an arc of M containing P and whose end points belong to K. Hence P belongs to the arc-curve $M(K)$.

COROLLARY 8A. *The set of all nodes of a compact continuous curve is a basic set of the continuous curve.*

With the use of the preceding theorem we may demonstrate the following three results.

THEOREM 9. *In order that a subset K of a compact continuous curve M, which is not cyclicly connected, be an irreducible basic set of M it is necessary and sufficient that* (1) *no point of K be a cut point of M, and* (2) *each point of K belong to some node of M and each node of M contain exactly one point of K.*

THEOREM 10. *The compact continuous curve M has a unique irreducible basic set if and only if every node of M is a point (i.e. every maximal cyclic curve of M contains at least two cut points of M). This unique irreducible basic set of M is the set of all end points of M.*

THEOREM 11. *The compact continuous curve M has a unique irreducible basic set if and only if the common part of every two basic sets of M is a basic set of M.*

THEOREM 12. *If x, y, z are three points of a continuous curve M, then either there is an arc of M containing all three points or there is a cyclic element C of M such that $M-C$ contains all three points x, y, z and no two of them lie in the same component of $M-C$.*

If the arc-curve $M(x+y)$ contains the point z, there is an arc of M with end points x and y and containing the point z. If not, then the component N of $M-M(x+y)$ containing z has just one limit point w in $M(x+y)$.* If the point w is either x or y, it is easy to see that the curve M contains an arc containing x, y and z. Suppose $x \neq w \neq y$. It is evident that neither x and z nor y and z can lie in the same component of M minus any subcontinuum of M containing w. If the point w is a cut point of the arc-curve $M(x+y)$, then w is a cut point of M† and is the desired cyclic element of M. If the point w is not a cut point of the arc-curve $M(x+y)$, it belongs to some simple closed curve J of $M(x+y)$, and let E be the maximal cyclic curve of M containing J. It is not difficult to see that the set E is a subset of the arc-curve $M(x+y)$. If E contains both x and y, then there exists an arc α of E with end points x and w and containing the point y. There exists an arc β with end points w and z and lying in N except for the point w. The set $\alpha+\beta$ is an arc of M containing x, y and z. If the curve E contains one of the two points x and y, but not both, we will suppose it contains x and not y. Let v be the limit point in E of the component H of $M(x+y)-E$ containing the point y. Since we have assumed that the point w was not a cut point of $M(x+y)$, the points w and v are distinct. There exists an arc γ with end points v and y and lying in $H+v$. There exists an arc η of E with end points w and v and containing the point x. Then $\beta+\gamma+\eta$ is an arc of M containing x, y and z. If the maximal cyclic curve E contains neither x nor y, since each point of E lies on some arc of M with end points x and y and no component of $M-E$ has more than one limit point in E, then x and y must belong to different components of $M-E$. Then E is the desired cyclic element of M.

COROLLARY 12A. *In order that every three points of a continuous curve M lie on some arc of M it is necessary and sufficient that $M-E$ should be connected or the sum of two connected sets, for each cyclic element E of M.*

THEOREM 13. *In order that a compact continuous curve M should have a basic set consisting of two points it is necessary and sufficient that every three points of M should lie on some arc of M.*

---

* See *Arc-curves, first paper*, Theorem 10, part (2).

† *Arc-curves, first paper*, Theorem 9, part (2).

The condition is necessary. If M contains two points x and y such that $M=M(x+y)$, then, by Theorems 3 and 4, the set $M-E$ is connected or the sum of two connected sets for each cyclic element E of M. Thus, by Corollary 12A, it follows that every three points of M lie on some arc of M.

The condition is sufficient. If M is cyclicly connected, any two points of M form a basic set of M. If M is not cyclicly connected, by Theorem 1 and a result due to G. T. Whyburn,* the set M has at least two nodes of itself. And the continuous curve M cannot have more than two nodes, for if it did contain three nodes, then, as each node contains a non-cut point of M, there would be an arc of M containing three such points by the condition of the theorem. But it has been pointed out that no non-cut point of M which belongs to some node of M can be interior to an arc of M whose end points do not belong to this node. Hence M contains just two nodes, and let x and y be points of these two nodes which are non-cut points of M. Then $x+y$ is a basic set of M by Theorem 8.

COROLLARY 13A. *If the compact subcontinuum K of the continuous curve M is the sum of a collection of cyclic elements of M, then K is a simple cyclic chain of M if and only if every three points of K lie on some arc of M.*

In view of Theorem 2 we get the following form for a theorem of G. T. Whyburn.†

THEOREM 14. *If x and y are points of a compact continuous curve M, then there exist two points w and z of M such that the arc-curve $M(w+z)$ contains the arc-curve $M(x+y)$ and is not a proper subset of the arc-curve with respect to M of any two points of M.*

In the next theorem we will give an analogue to Theorem 14 for a more general arc-curve. Before doing this we will prove several lemmas.

LEMMA 15A. *A continuous curve M contains a countable set of points $x_1, y_1, x_2, y_2, x_3, y_3, \cdots$ such that* (1) *every point of M is either an end point of M or a point of one of the arc-curves $M(x_i+y_i)$,* (2) *for any positive integer n, the set $M(x_1+y_1)+M(x_2+y_2)+\cdots+M(x_n+y_n)$ is a continuous curve M_n,* (3) *M_n has only one point in common with the arc-curve $M(x_{n+1}+y_{n+1})$ and this is either the point x_{n+1} or the point y_{n+1},* (4) *for each positive number ϵ and each hypersphere S there exists an integer ρ such that for $n>\rho$, the set $M-M_n$ contains no component of diameter greater than ϵ which contains a point interior to S.*

---

* See *Structure*, Theorem 14.

† See *Structure*, Theorem 15.

This lemma may be proved in much the same way as a theorem of R. L. Wilder of which it is an analogue.*

LEMMA 15B.† *If K is an arc-curve of the continuous curve M (or if K is a connected collection of cyclic elements of M), then there are only a countable number of nodes of K that are not nodes of M.*

Let G be the collection of all nodes of K that are not nodes of M. Since every cyclic element of K is a cyclic element of M and no continuous curve contains more than a countable number of maximal cyclic curves, all but a countable number of the nodes of the collection G are end points of K. Let G' be the collection of all elements of G that are end points of K. As each cyclic element of K is a cyclic element of M and no node in G is a node of M, each point of G' is a cut point of M. Then each point of G' belongs to one of the arc-curves $M(x_i+y_i)$. If any arc-curve $M(x_i+y_i)$ contains three points of G', then, by Theorem 13,‡ there is an arc α of $M(x_i+y_i)$ with two of these points as end points and containing the third. The arc α belongs to the arc-curve K since every arc-curve contains every arc of M whose end points belong to the arc-curve. But no end point of an arc-curve K is interior to an arc of K. Then no arc-curve $M(x_i+y_i)$ contains more than two points of the collection G'. Then the collection G' is countable. Hence the collection G must be countable.

In proving the last lemma we have established a somewhat more general result.

LEMMA 15C. *If K is an arc-curve of the continuous curve M, then there are only a countable number of nodes of K that are not end points of M.*

THEOREM 15. *If K is any subset of the compact continuous curve M, there exists a subset K' of M such that (1) the arc-curve $M(K')$ is closed and contains the arc-curve $M(K)$, (2) if K'' is any subset of M such that the arc-curve $M(K')$ is a subset of the arc-curve $M(K'')$ then every node of $M(K')$ is a node of $M(K'')$, (3) the set K' is countable unless $M(\overline{K})$ contains an uncountable number of end points of M, (4) if $M(\overline{K})$ has n nodes (n finite and greater than one) then K' contains just n points.*

If $M(\overline{K})$ is a point or is cyclicly connected, we may apply Theorem 14 and obtain a set $K'=w+z$ such that the arc-curve $M(w+z)$ contains the

---

* R. L. Wilder, *Concerning continuous curves*, loc. cit., Theorem 15.

† This lemma is closely related to a theorem of R. L. Moore. See his paper *Concerning the cut points of continuous curves and of other closed and connected point sets*, loc. cit., Theorem B*.

‡ Compactness was not used in proving the necessity of Theorem 13.

arc-curve $M(K)$, and is not a proper subset of any arc-curve of two points of M. It is not difficult to see that the set K' satisfies the conditions of the theorem.

If $M(\overline{K})$ is not cyclicly connected, let G be the set of all nodes of $M(\overline{K})$. Let E be any node of G. If E is a node of M, let P_e be a non-cut point of M which belongs to E. If E is not a node of M, then E is either a cut point of M or a maximal cyclic curve of M containing more than one cut point of M. In either case there is a component H_e of $M-M(\overline{K})$ that has just one limit point Q_e in E, and Q_e is not a cut point of $M(\overline{K})$. As in the proof of Theorem 8 we may show that H_e contains a node E' of M. Let $P_{e'}$ be a non-cut point of M which belongs to the node E'. Let K' be the set of all points P_e and all the points $P_{e'}$. We shall show that the set K' has the desired properties.

By Theorem 8 the set $[P_e]+[Q_e]$ is a basic set of $M(\overline{K})$. Hence if each point of this set belongs to the arc-curve $M(K')$, then every point of $M(\overline{K})$ belongs to $M(K')$ since every other point of $M(\overline{K})$ is on an arc whose end points belong to the set $[P_e]+[Q_e]$.* Since each point P_e belongs to K', it belongs to $M(K')$. Let q_e^1 be a point of $[Q_e]$ and let $P_{e'}^1$ be the corresponding point of $[P_{e'}]$ and H_e^1 be the component of $M-M(\overline{K})$ containing $P_{e'}^1$. Let P_e^2 be any point of $K'-P_{e'}^1$. Then $M-H_e^1$ contains P_e^2 and thus any arc of M from $P_{e'}^1$ to P_e^2 must contain the point Q_e^1. Hence the arc-curve $M(K')$ contains the arc-curve $M(\overline{K})$.

We shall now show that $M(K')$ is closed. If $M(K')$ is not closed, every point of $\overline{M(K')}-M(K')$ is a limit point of K'.† Let X be any point of $\overline{M(K')}-M(K')$. Either (1) the point X is a limit point of points of K' that belong to $M-M(\overline{K})$, or (2) it is a limit point of points of K' that belong to $M(\overline{K})$. Since $M(\overline{K})$ is a continuous curve‡ and is a subset of $M(K')$, case (2) is impossible. In case (1) the point X is a limit point of the subset $[P_{e'}]$ of K'. But no component of $M-M(\overline{K})$ contains more than one point of $[P_{e'}]$ and thus every limit point of $[P_{e'}]$ belongs to $M(\overline{K})$. But X cannot be such a point for $M(\overline{K})$ is a subset of $M(K')$ while X belongs to $\overline{M(K')}-M(K')$. Thus we obtain a contradiction in both cases and the set $M(K')$ must be closed.

Every point P_e of K' is a non-cut point of M which belongs to a node E of M and every point $P_{e'}$ of K' is a non-cut point of M that belongs to some node E' of M. It follows easily that $M(K')$ contains every point of all the nodes $[E]$ and $[E']$ of M. Then each of these sets is a node of $M(K')$ and of every arc-curve of M containing $M(K')$. This proves part (2).

* *Arc-curves, first paper*, Theorem 3.
† *Arc-curves, first paper*, Theorem 2.
‡ *Arc-curves, first paper*, Theorem 7.

From the definition of the set K' it is seen that the cardinal number of the set K' is the same as that of the set of nodes of the arc-curve $M(\overline{K})$, except when $M(\overline{K})$ is a point or is cyclicly connected. This establishes part (4) and, with Lemma 15C, establishes part (3).

COROLLARY 15A. *If the connected subset H of the compact continuous curve M is the sum of a collection of cyclic elements of M, then there exists a subcontinuum K of M such that* (1) *K contains H and is the sum of a collection of cyclic elements of M,* (2) *if N is any connected subset of M which contains K and is the sum of a collection of cyclic elements of M, then every node of K is a node of N,* (3) *the set of nodes of K has the same cardinal number as the set of nodes of $\overline{H}$, unless $\overline{H}$ is a point or is cyclicly connected, in which case K may have one or two nodes.*

IV. THE IRREDUCIBLE BASIC SETS AND THE SET OF END POINTS

In *Arc-curves, first paper* it was shown that every basic set of a continuous curve contains every end point of the continuous curve. However, a basic set may contain many other points, in fact it may be the entire continuous curve. The set of end points may or may not be a basic set of the curve. The condition under which this was true was given in Theorem 10. If the set of end points is a basic set, it is an irreducible basic set. An irreducible basic set of a continuous curve may contain a countable number of points that are not end points of the curve,* but in many respects an irreducible basic set is similar to the set of all end points of the continuous curve. No end point of a continuous curve is a cut point of the curve, and similarly no point of an irreducible basic set is a cut point of the curve. No end point of a continuous curve is interior to an arc of the curve. We have the somewhat analogous property that no point P of an irreducible basic set of a continuous curve M is interior to any arc of M, neither of whose end points belong to that node of M containing the point P. Karl Menger,† G. T. Whyburn‡ and the author§ have shown that the set of all end points

* By Theorem 9 there is a (1, 1) correspondence between the points of an irreducible basic set and the set of nodes in which they are contained. By Lemma 15C, all but a countable number of nodes are end points. Hence all but a countable number of the points of an irreducible basic set are end points of the continuous curve.

† K. Menger, *Grundzüge einer Theorie der Kurven*, Mathematische Annalen, vol. 95 (1925), pp. 277–306. Menger defines the term *end point* in a different sense from that of R. L. Wilder, but the two definitions have been shown to be equivalent by H. M. Gehman. See his paper *Concerning end points* etc., loc. cit., Theorem 1.

‡ G. T. Whyburn, *Concerning continua in the plane*, loc. cit., Theorem 21.

§ W. L. Ayres, *Concerning continuous curves and correspondences*, loc. cit., Theorem 5.

of a continuous curve is totally disconnected.* It has been shown by the author† that if K is any subset of the set of end points of a continuous curve M, then $M-K$ is strongly connected.‡ We shall show that both of these properties hold for irreducible basic sets of a continuous curve.

THEOREM 16. *If M is a continuous curve, then every irreducible basic set of M is totally disconnected.*

Let K be an irreducible basic set of M and suppose that K contains a connected subset H containing more than one point. Since the set of end points of M is totally disconnected, the set H contains a point P which is not an end point of M. Since no point of K is a cut point of M,§ the point P belongs to some simple closed curve J of M and let C be the maximal cyclic curve of M containing J. By a theorem of G. T. Whyburn,‖ the common part of H and C is a connected set N. As P is not a cut point of M, the set N contains points other than the point P. Then the irreducible basic set K contains a connected set N containing more than one point and lying in a maximal cyclic curve C of M. But this is impossible for no maximal cyclic curve of a continuous curve contains more than two points of an irreducible basic set of the continuous curve by Theorem 9.

THEOREM 17. *If K is any subset of an irreducible basic set of a continuous curve M, which is not cyclicly connected, then $M-K$ is arc-wise connected.*

Let P_1 and P_2 be points of $M-K$. There are four possible cases: (1) both P_1 and P_2 belong to a maximal cyclic curve C of M which is a node, (2) neither P_1 nor P_2 belongs to a node of M containing a point of K, (3) both P_1 and P_2 belong to nodes of M containing points of K but not both to the same node, (4) either P_1 or P_2 belongs to a node of M containing a point of K but the other does not. In case (1), since M is not cyclicly connected, the curve C contains at most one point Q of K by Theorem 9. And as Q is not a cut point of C, there is an arc of $C-Q$ whose end points are P_1 and P_2. Since this arc belongs to $C-Q$, it belongs to $M-K$. In case (2) let α be any arc of M with end points P_1 and P_2. Every point of K is a non-cut point of M belonging to some node of M and no such point is interior to any arc of M whose end points do not belong to the node. Hence no point of K is an interior point of the arc α.

* A point set is said to be *totally disconnected* if it contains no connected subset containing more than one point.

† *Concerning continuous curves and correspondences*, loc. cit., Theorem 6.

‡ A point set K is said to be *strongly connected* if each pair of points of K belongs to some continuum which is a subset of K.

§ *Arc-curves, first paper*, Theorem 13.

‖ *Structure*, Theorem 30.

In case (3) let N_1 and N_2 be the nodes of M containing P_1 and P_2 respectively. Let Q_1 and Q_2 be the points of K and X_1 and X_2 be the cut points of M belonging to N_1 and N_2 respectively. Evidently $P_i \neq Q_i \neq X_i$ $(i=1, 2)$. Then, as $N_i - Q_i$ is connected, there exists an arc α_i of $N_i - Q_i$ from P_i to X_i. Let α_3 be any arc of M with end points X_1 and X_2. It is seen that $M - N_1 - N_2$ contains the set $<\alpha_3>$ and that no point of K belongs to the arc α_3. Then the set $\alpha_1 + \alpha_2 + \alpha_3$ is an arc of M with end points P_1 and P_2 and containing no point of K.

Case (4) is very similar to case (3) and need not be considered separately.

COROLLARY 17A. *If K is an irreducible basic set of a continuous curve M, then $M - K$ consists of a finite number of components.*

LEMMA 18A. *In order that a connected subset K of a continuous curve M be connected im kleinen at a point P of K it is necessary and sufficient that the set $H \cdot K$ should be connected im kleinen at the point P for each maximal cyclic curve H of M containing the point P.*

The condition is necessary. If P belongs to no maximal cyclic curve of M, the condition is satisfied vacuously. If not, let H be any maximal cyclic curve of M containing the point P, and let ϵ be any positive number. Since K is connected im kleinen at P, there exists a positive number δ_ϵ such that any point of K whose distance from P is less than δ_ϵ lies with P in a connected subset of K of diameter less than ϵ. Let Q be any point of $H \cdot K$ such that $d(P, Q) < \delta_\epsilon$ and let N be a connected subset of K of diameter less than ϵ containing both P and Q. As H is a maximal cyclic curve of M, the set $N \cdot H$ is connected. It is a subset of $H \cdot K$ of diameter less than ϵ and contains P and Q. Hence $H \cdot K$ is connected im kleinen at P.

The condition is sufficient. If ϵ is any positive number, there are only a finite number of components of $M - P$ of diameter greater than ϵ,* and let $N_1, N_2, N_3, \cdots, N_m$ denote these components. If P does not belong to a simple closed curve of $N_i + P$, then there is a point Q_i of N_i such that every point of the component L_i of $N_i + P - Q_i$ containing the point P is within a

* For two dimensions this result may be found in my paper *Note on a theorem concerning continuous curves*, Annals of Mathematics, (2), vol. 28 (1927), pp. 501–2. While this result does not hold in space of more than two dimensions, it may be shown that *if M is a continuous curve and N is a compact subcontinuum of M, ϵ is any positive number and η is any positive number less than ϵ, then only a finite number of the components C of $M - N$ that are of diameter greater than ϵ are such that the set of limit points of C in N is of diameter less than η.* The result of the paper mentioned above may be obtained from this theorem since if M lies in the plane and N is a continuous curve, we can show that there are only a finite number of components C of $M - N$ such that the set of limit points of C in N is of diameter greater than or equal to η.

distance ϵ of P.* It is easy to see that the set $K\cdot(L_i+Q_i)$ is connected. Let $\delta_{\epsilon i}=\frac{1}{2}d(P, N_i+P-L_i)$. Then every point of $K\cdot N_i$ whose distance from P is less than $\delta_{\epsilon i}$ belongs to $K\cdot(L_i+Q_i)$ and thus lies with P in a connected subset of K of diameter less than ϵ.

If P belongs to a simple closed curve of N_i+P, there is one (and only one) maximal cyclic curve C_i of N_i+P containing P. The set C_i is also a maximal cyclic curve of M. Hence by the condition there exists a positive number $\delta_{\epsilon i}'$ such that any point of $K\cdot C_i$ whose distance from P is less than $\delta_{\epsilon i}'$ lies with P in a connected subset of $K\cdot C_i$ every point of which is within a distance $\epsilon/2$ of P. Since only a finite number of the components of N_i+P-C_i are of diameter greater than a given positive number and each of them has just one limit point in C_i,† there exists a positive number $\delta_{\epsilon i}''$ such that if X is any point of N_i+P-C_i whose distance from P is less than $\delta_{\epsilon i}''$, then X belongs to a component of diameter less than $\epsilon/2$ and the limit point of this component in C_i is within a distance $\delta_{\epsilon i}'$ of P. Let $\delta_{\epsilon i}$ be the smaller of the numbers $\delta_{\epsilon i}'$ and $\delta_{\epsilon i}''$. Now let Q be any point of $K\cdot N_i$ whose distance from P is less than $\delta_{\epsilon i}$. If Q belongs to N_i+P-C_i, let L_i be the component of N_i+P-C_i containing Q and let Y_i be its limit point in C_i. The set $W_i=K\cdot L_i+Y_i$ is a connected subset of K of diameter less than $\epsilon/2$ and containing the point Y_i. If Q belongs to C_i, let W_i and Y_i be identical with the point Q. Since $d(P, Y_i)<\delta_{\epsilon i}'$, there is a connected subset V_i of $K\cdot C_i$ containing P and Y_i, every point of which is within a distance $\epsilon/2$ of P. Then every point Q of $K\cdot N_i$ whose distance from P is less than $\delta_{\epsilon i}$ lies with P in a connected subset W_i+V_i of $K\cdot N_i$ every point of which is within a distance ϵ of P.

If δ_ϵ is the smallest of the numbers $\delta_{\epsilon 1}, \delta_{\epsilon 2}, \delta_{\epsilon 3}, \cdots, \delta_{\epsilon m}$, it is easily seen that every point of K whose distance from P is less than δ_ϵ lies with P in a connected subset of K every point of which is within a distance ϵ of P.

THEOREM 18. *If K is any subset of an irreducible basic set of a continuous curve M, then $M-K$ is connected im kleinen.*

If M is cyclicly connected, then K has at most two points and $M-K$ is connected im kleinen since it is an open subset of M. If M is not cyclicly connected, by Theorem 17 the set $M-K$ is connected. Then, by the preceding lemma, the theorem is true if $(M-K)\cdot C$ is connected im kleinen for every maximal cyclic curve C of M. As M is not cyclicly connected, the maximal

* Any such point P is an end point of N_i+P by Theorem 3 of my paper *Concerning continuous curves and correspondences*, loc. cit., and thus the point Q_i exists. See H. M. Gehman, *Concerning end points* etc., loc. cit., especially property 7.

† If η is chosen as any positive number less than ϵ, then there is no component C of N_i+P-C_i such that the set of limit points of C in C_i is of diameter greater than η.

cyclic curve C contains at most one point of K. It is evident that any maximal cyclic curve or a maximal cyclic curve minus one point is connected im kleinen.

Whyburn* has defined the end elements of a continuous curve and has shown that in many respects they resemble the end points of a continuous curve. He has shown that if the continuous curve is neither cyclicly connected nor the sum of two cyclicly connected continuous curves, the set of end elements is totally disconnected in the sense that it contains no connected subset containing more than one end element. However the end-point properties mentioned in the two preceding theorems of this paper are not true for end elements, as is seen in the following example. Let M consist of the square with vertices $(0, 0)$, $(1, 0)$, $(1, -1)$ and $(0, -1)$ together with the circles with center $(1/n, r_n)$ and radius r_n where $n=1, 2, 3, \cdots$, and $r_n=1/(2n(n+1))$. In this example each circle is an end element of the continuous curve M. If K is the set of end elements, then $M-K$ consists of a countable infinity of components and is not connected im kleinen at the point $(0,0)$.

---

* See *Structure*, p. 178 and Theorems 16–19.

THE UNIVERSITY OF TEXAS,
AUSTIN, TEXAS

ON FERMAT'S LAST THEOREM*

BY

H. S. VANDIVER

In two recent papers† the writer stated without proof four theorems concerning Fermat's last theorem. These results were employed to prove the theorem for all exponents greater than 2 and less than 211. In the present paper the proofs of Theorems I to IV of my previous papers will be given, as well as one of an additional Theorem V.

I shall first indicate the relation of the present results to the previous work done on the problem. In the year 1850 Kummer‡ proved that the equation

$$\alpha^l + \beta^l + \gamma^l = 0 \tag{1}$$

is impossible if α, β and γ are integers in the field $k(\zeta)$ prime to each other; $\zeta = e^{2i\pi/l}$ and l is an odd prime greater than 3 such that $B_1, B_2, \cdots, B_{(l-3)/2}$ have numerators which are prime to l, where $B_1 = 1/6$, $B_2 = 1/30$, etc. are the Bernoulli numbers.

As shown by Kummer§ the above condition concerning the Bernoulli numbers is equivalent to the statement that the class number of the field $k(\zeta)$ is prime to l. Primes l which have this property are called regular. In particular he showed that the second factor¶ of the class number is divisible by l only if the first factor of the class number is divisible by l. This is not a sufficient condition, however, for the divisibility of the second factor by l. He computed‖ the first factor of the class number for all primes $l < 164$. He found that l is regular for all these primes excepting $l = 37, 59, 67, 101, 103, 131, 149,$ and 157.

For $l = 157$, the first factor of the class number is divisible by 157^2 and not by 157^3, but in the other cases by the first power of l only. For $l = 101$,

* Presented to the Society, December 27, 1928; received by the editors in April, 1929.

† Proceedings of the National Academy of Sciences, vol. 15 (1929), pp. 48; 109–110.

‡ Crelle's Journal, vol. 40 (1850), pp. 130–138. Proof extended by Hilbert, *Algebraische Zahlkörper*, Jahresbericht der Deutschen Mathematiker-Vereinigung, 1894, pp. 517–523.

§ Crelle's Journal, vol. 40 (1850), pp. 117–129; Journal de Mathématiques, (1), vol. 16 (1851), pp. 473–486; abstract in Berlin Monatsberichte, 1847, pp. 305–319. Also Vandiver, Bulletin of the American Mathematical Society, vol. 25 (1918–19), pp. 458–461. For a full history of the problem as far as it is connected with algebraic numbers, cf. Bulletin of the National Research Council, No. 62, February, 1928.

¶ Crelle's Journal, vol. 40 (1850), pp. 93–116; Hilbert, Bericht, p. 377.

‖ Crelle's Journal, vol. 40 (1850), p. 117; Journal de Mathématiques, (1), vol. 16 (1851), p. 473. Berlin Monatsberichte, 1847, p. 319. Berlin Monatsberichte, 1874, pp. 239–248.

BOSTON UNIVERSITY
COLLEGE OF LIBERAL ARTS
LIBRARY

the 34th Bernoulli number is divisible by l; for $l=103$, the 12th Bernoulli number; for $l=131$, the 11th Bernoulli number; for $l=149$, the 65th Bernoulli number; for $l=157$, the 31st and 55th Bernoulli numbers.

In the year 1857 Kummer attempted to prove that

$$x^l + y^l + z^l = 0 \tag{2}$$

is impossible for rational integers x, y and z under the following three assumptions:

ASSUMPTION I. The first factor of the class number of $k(\zeta)$ is divisible by l but not by l^2.

ASSUMPTION II. If $B_n \equiv 0 \pmod{l}$, $n<(l-1)/2$, there exists an ideal in $k(\zeta)$ with respect to which as a modulus the unit

$$E_n = \prod_{i=0}^{(l-3)/2} \epsilon(\zeta^{r^i})^{r^{2in}}$$

is not congruent to the lth power of an integer in $k(\zeta)$. Here

$$\epsilon = \left(\frac{(1-\zeta^r)(1-\zeta^{-r})}{(1-\zeta)(1-\zeta^{-1})}\right)^{1/2}$$

and $\epsilon(\zeta^{r^i})$ indicates the unit obtained from ϵ by the substitution (ζ/ζ^{r^i}); also r is a primitive root of l.

ASSUMPTION III. The Bernoulli number B_{nl} is not divisible by l^3.

He then applied these criteria to the exponents $l=37$, 59, and 67, the only l's less than 100 which are not regular. From the remarks made above concerning the computations of the first factors of the class numbers it follows that Assumption I is satisfied, and carrying out extensive computations Kummer concluded that Assumptions II and III were satisfied also for the three values of l mentioned above. He gave, however, only the results of the latter computations and in connection with Assumption III did not give the formula that he used on which his calculations were based.

If in (2) x, y and z be prime to l this is called the first case of Fermat's last theorem; if x, y and z are prime to each other and $xyz \equiv 0 \pmod{l}$ then this is called case II of the theorem. *Before the appearance of the papers of Kummer and between their appearance and the present time many contributions have been made to the first case of the theorem; in particular* the theorem has been*

* Dickson, Messenger of Mathematics, (2), vol. 38 (1908), pp. 14–32; Quarterly Journal of Mathematics, vol. 40 (1908), pp. 27-45; Beeger, Messenger of Mathematics, vol. 55 (1925), pp. 17–26.

proved for all exponents $2<l<14{,}000$ *in case* I. *The situation in connection with the second case is quite different, however. With the exception of the two articles by the writer cited above and another of Bernstein's* no paper has been published since the year* 1857 *which gives any new criteria for the solution of* (2) *in case* II *which are independent of* x, y *and* z *and expressible in terms of rational integers.*

In the year 1920 the writer† pointed out that the proofs of the results given in Kummer's 1857 Memoir on Fermat's last theorem are inaccurate and incomplete in several respects.

He later‡ completed Kummer's proofs by modifications and extensions of the latter's arguments.

As noted above the prime $l=157$ is irregular and the first factor of the class number $k(\zeta)$ for this case is divisible by l^2. Hence the methods of Kummer's 1857 paper do not apply to this exponent since the first assumption states that the first factor of the class number $k(\zeta)$ is not divisible by l^2. *The present paper is the result, in part, of efforts to obtain criteria for the second case of Fermat's last theorem which would yield a proof of the theorem for the case* $l=157$ *at least, and hence to effect an advance over Kummer's work. Theorems* I *and* IV *as proved in the present paper each yield a proof for this exponent as well as the proofs for all other irregular primes* l *as exponents which are less than* 211.

Concerning the computations which prove Fermat's last theorem in the second case for all exponents greater than 2 and less than 211, all primes less than this limit were tested as to being irregular. For a particular l, this was accomplished by testing each of the Bernoulli numbers $B_1, B_2, \cdots, B_{(l-3)/2}$ as to the divisibility of their numerators by l. By a systematic method, for primes l less than 100 the necessary computations were made by Mrs. A. C. S. Williams; for the primes l between 100 and 200 they were made by Professor Elizabeth T. Stafford. The results showed that the only irregular primes in this range were those that had been discovered by Kummer, who

* Bernstein, Göttinger Nachrichten, Mathematisch-Physikalische Klasse, 1910, pp. 507–516. It was pointed out by the writer, however (Proceedings of the National Academy of Sciences, vol. 6 (1920), pp. 416–421), that Bernstein's first theorem constitutes no advance over the results of Kummer's paper of the year 1850 on Fermat's last theorem, and it was shown by Pollaczek (Mathematische Zeitschrift, vol. 21, pp. 36–38) that in order for Bernstein's second theorem to yield a proof of Fermat's last theorem for a particular exponent l a number of conditions concerning the field $k(\zeta)$ must be satisfied. In particular, his theorem yields no proof for any of the three cases $l=37$, 59 and 67.

† Proceedings of the National Academy of Sciences, vol. 6 (1920), pp. 266–269.

‡ Bulletin of the American Mathematical Society, vol. 28 (1922), pp. 400–407; Proceedings of the National Academy of Sciences, vol. 12 (1926), pp. 767–772.

carried his computations up to and including $l=163$. These computations concerning the Bernoulli numbers obviously do not give as much information as Kummer obtained, as he, in his work, computed the actual values of the first factor of the class number of each l. However, I regard the former computations as a sufficient check on Kummer's results.

The computations in connection with the irregular primes in testing the criteria of Theorems I to IV of this paper were carried out by Professor E. T. Stafford, Mrs. A. C. S. Williams, Mr. S. S. Wilks, and the writer.

In none of the criteria found by Kummer for the second case of Fermat's last theorem was any use made of the fact that x, y and z in (2) are rational integers. The criteria apply also if x, y and z are integers in the field defined by $(\zeta+\zeta^{-1})$, prime to each other. In Theorem V of the present article I give for the first time criteria for the second case of the theorem which are obtained by use of the fact that x, y and z are rational.

1. We first consider

LEMMA 1. *If the ideal* (ω) *in* $k(\zeta)$ *is the* l*th power of an ideal in that field and* ω *is a primary number, then* (ω) *is the* l*th power of a principal ideal in the field provided that the second factor of the class number of* $k(\zeta)$ *is prime to* l.

This was proved in full in another paper by the writer.* It is also noted that this lemma may also be proved using a result of Takagi.† This follows since Takagi shows that if the second factor of the class number of $k(\zeta)$ is prime to l then any singular primary number in the field may be expressed in the form $\eta\theta^l$ where η is a unit and θ a number in $k(\zeta)$. This statement is equivalent to the one in the lemma since a singular primary number ω is defined to be such that (ω) is the lth power of an ideal which is not a principal ideal, ω being a number in $k(\zeta)$.

We now proceed to

LEMMA 2. *If a unit* η *exists in* $k(\zeta)$ *which is of the form*

$$\eta \equiv c^l \pmod{\lambda^{2l}}$$

where c *is a rational integer, then* η *is the* l*th power of a unit in* $k(\zeta)$ *except possibly when*

$$B_{nl} \equiv 0 \pmod{l^3}$$

for n *some integer in the set* $1, 2, \cdots, (l-3)/2$; $\lambda=(1-\zeta)$.

To prove this we set

* Proceedings of the National Academy of Sciences, 1929. Note that the relation just above (7) gives $\omega\omega_{-1}^{l-1}=\theta^l$ which with $\omega\omega_{-1}=\gamma_1\sigma_1{}^l$ yields $\omega=\gamma\sigma^l$, where γ and γ_1 are units.

† Crelle's Journal, vol. 157 (1927), p. 236. Cf. also Bernstein, Göttinger Nachrichten, loc. cit., p. 514.

$$\eta^t = E_1^{a_1} E_2^{a_2} \cdots E_{(l-3)/2}^{a_{(l-3)/2}} \tag{3}$$

the a's and t being rational integers Since the E's form an independent system in $k(\zeta)$ this is always possible. Further we may assume that l is not a common factor of all the integers $t, a_1, a_2, \cdots, a_{l_1}$, $l_1=(l-3)/2$. For if it is, and $t=l^h t_1$, $a_1=l^h a_1^1, \cdots, a_{l_1}=l^h a_{l_1}^1$ with l prime to at least one of the integers $t_1, a_1^1, a_2^1, \cdots, a_{l_1}^1$, then (3) holds with t_1 in lieu of t, a_1^1 in lieu of a_1, etc., since all the E's are real.

We may, by the hypothesis in the theorem, set

$$\eta = c^l + \lambda^{2l}\alpha_1 = c^l + l^2\alpha, \tag{3a}$$

where α_1 and α are integers in $k(\zeta)$. We may write

$$\alpha = m_0 + m_1\zeta + \cdots + m_{l-2}\zeta^{l-2},$$

with the m's rational integers. If w is an arbitrary variable, set

$$\alpha(w) = m_0 + m_1 w + \cdots + m_{l-2}w^{l-2}.$$

Now ϵ may be put in the form

$$\zeta^{(1-r)/2}\frac{\zeta^r - 1}{\zeta - 1}.$$

Set $r' \equiv (1-r)/2 \pmod{l}$, with r' a positive rational integer. Put

$$\epsilon(w) = w^{r'}\frac{w^r - 1}{w - 1}$$

and

$$E_i(w) = \prod_{j=1}^{l_1} \epsilon(w^{r^j})^{r^{-2ij}};$$

$\epsilon(w)$ is obviously a polynomial in w with rational integral coefficients. Set

$$\rho = r^{l^2};$$

then $(E_i(w))^\rho$ is also a polynomial in w with rational integral coefficients, since it may be written as

$$\prod_{j=0}^{l_1} \epsilon(w^{r^j})^{\rho r^{-2ij}}$$

and ρr^{-2ij} is a positive integer because $l^2 > 2ij$. Now consider the expression

$$A = \prod_{i=1}^{l_1} E_i^{a_i \rho^{(l-1)}}(w) - (c^l + l^2\alpha(w))^{\rho^{(l-1)}t}.$$

It is a polynomial in w with rational integral coefficients, and vanishes for $w=\zeta$ by (3) and (3a).

Since

$$\frac{w^l-1}{w-1}$$

is irreducible in the rational field it therefore follows that

$$A = X\frac{w^l-1}{w-1},$$

where X is a polynomial in w with rational integral coefficients. Further if A is divided by w^l-1 we obtain a remainder which is of degree $<l$ and divisible by

$$\frac{w^l-1}{w-1}.$$

Hence this remainder is of the form

$$\frac{b(w^l-1)}{w-1}$$

where b is a rational integer. We therefore write

$$\text{(3b)}\qquad \sum_{i=1}^{h} E_i^{a_i\rho^{(l-1)}}(w) = (c^l + l^2\alpha(w))^{\rho^{l(l-1)}} + V(w^l-1) + b\frac{w^l-1}{w-1},$$

where V is a polynomial in w with rational integral coefficients. Put $w=1$ in the above; we have

$$\text{(3c)}\qquad \sum_{i=1}^{l} r^{(l-1)a_i\rho(1+r^{-2i}+\cdots+r^{-(l-3)i})} = c^{l\rho^{(l-1)l}} + bl.$$

Since

$$1 + r^{-2i} + \cdots + r^{-(l-3)i} = \frac{r^{-i(l-1)}-1}{r^{-2i}-1},$$

then

$$\rho\frac{r^{-i(l-1)}-1}{r^{-2i}-1}$$

is an integer divisible by l, and since

$$c^{l(l-1)} \equiv r^{l(l-1)} \equiv 1 \pmod{l^2},$$

the relation (3c) gives

$$bl \equiv 0 \pmod{l^2}; \qquad b \equiv 0 \pmod{l}.$$

Set $b=b_1 l$ and $w=e^v$; then (3b) gives

$$(3d)\qquad \prod_{i=1}^{h} E_i^{a_i\rho^{(l-1)}(e^v)} = (c^l + l^2\alpha(e^v))^{\rho^{(l-1)t}} + V(e^{vl} - 1) + b_1 l\,\frac{e^{vl} - 1}{e^v - 1}\,.$$

To proceed further with this relation we first derive some properties of the E's. We have

$$\frac{d}{dv}\log\frac{e^{rv} - 1}{e^v - 1} = \frac{r - 1}{2} + (r^2 - 1)\frac{B_1}{2!}v - (r^4 - 1)\frac{B_2}{4!}v^3 + \cdots,$$

$$B_1 = 1/6,\ B_2 = 1/30, \cdots;\ |v| < \frac{2\pi}{r},$$

whence

$$\left[\frac{d^{2i}\log\dfrac{e^{rv} - 1}{e^v - 1}}{dv^{2i}}\right]_{v=0} = (-1)^{i+1}\,\frac{B_i}{2i}(r^{2i} - 1).$$

We shall now show that

$$(3e)\qquad \left[\frac{d^k \log F(e^{rv})}{dv^k}\right]_{v=0} = r^k\left[\frac{d^k \log F(e^v)}{dv^k}\right]_{v=0}$$

where $F(e^v)$ is a polynomial in e^v. Set

$$\frac{d^{k-1}\log F(e^v)}{dv^{k-1}} = \frac{A(e^v)}{B(e^v)}$$

where $A(e^v)$ and $B(e^v)$ are polynomials in e^v. Assume that

$$\frac{d^{k-1}\log F(e^{rv})}{dv^{k-1}} = r^{k-1}\frac{A(e^{rv})}{B(e^{rv})}\,.$$

This is true for $k=2$. We shall then show that this then holds for k in lieu of $k-1$. Differentiate each member of this equation and we obtain

$$\frac{d^k \log F(e^{rv})}{dv^k} = \frac{r^k}{B(e^{rv})}\,\frac{dA(e^{rv})}{dv} - \frac{r^k A(e^{rv})}{B^2(e^{rv})}\,\frac{dB(e^{rv})}{dv} = r^k G(e^{rv})$$

and in the same way

$$\frac{d^k \log F(e^v)}{dv^k} = G(e^v).$$

Putting $v=0$ in these relations we obtain (3e) which gives the relation desired. Applying this to the E's,

$$(4)\qquad \left[\frac{d^{2i}\log E_n(e^v)}{dv^{2i}}\right]_{v=0}$$
$$= (1 + r^{2i-2n} + r^{4i-4n} + \cdots r^{2i(l-3)/2 - 2n(l-3)/2})\frac{(-1)^{i+1}B_i}{2i}(r^{2i} - 1).$$

Hence if $i \not\equiv n$,

$$\left[\frac{d^{2i}\log E_n(e^v)}{dv^{2i}}\right]_{v=0} = \frac{r^{(l-1)(i-n)}-1}{r^{2i-2n}-1}(-1)^{i+n}\frac{B_i}{2i}(r^{2i}-1), \tag{4a}$$

and if $i=n$,

$$\left[\frac{d^{2n}\log E_n(e^v)}{dv^{2n}}\right]_{v=0} = \frac{(-1)^{n+1}B_n(l-1)(r^{2n}-1)}{4n}. \tag{5}$$

So far in this paper we have put no limitation on r except that it be a primitive root of l; we may then further assume that $r^{l-1}\equiv 1 \pmod{l^2}$. In (3d) take logarithms of both members, differentiate $2kl$ times, $k=1, 2, \cdots, l_1$, with respect to v, set $v=0$ and reduce the result modulo l^2. In the left hand member we employ (4) and (4a) modulo l^2 and using $r^{l-1}\equiv 1 \pmod{l^2}$ we obtain, modulo l^2,

$$a_k r^{l^2}\frac{(l-1)^2}{2}\quad\frac{(-1)^{kl+1}B_{kl}}{2kl}(r^{2kl}-1).$$

As to the right hand member, denote it by Z; then

$$\frac{d^{kl}\log Z}{dv} = \frac{d^{kl-1}}{dv}\left(\frac{1}{Z}\frac{dZ}{dv}\right) = D_{kl-1}(Z^{-1}D(Z)),$$

say. With the latter notation we have

$$D_{kl-1}(Z^{-1}D(Z)) = Z^{-1}D_{kl}(Z) + (kl-1)D_{kl-1}(Z)D(Z^{-1}) + \cdots + D(Z)D_{kl-1}(Z^{-1}).$$

When we substitute $v=0$ in this relation we obtain on the right hand side the sum of a number of fractions each of whose denominators is of the form 1 (mod l^2) and whose numerators are each divisible by l^2 except possibly the term $[Z^{-1}D_{kl}(z)]_{v=0}$. It may be shown that this term is divisible by l^2, however, since

$$D_{kl}\left[(e^{vl}-1)V + b_1 l\frac{e^{vl}-1}{e^v-1}\right]_{v=0} \equiv [(e^{vl}-1)D_{kl}(V) + klD_{kl-1}(V)D(e^{vl}-1)]_{v=0} \equiv 0 \pmod{l^2},$$

observing that

$$D_i\left(\frac{e^{vl}-1}{e^v-1}\right) \equiv 0 \pmod{l};\quad i \not\equiv 0 \pmod{(l-1)},$$

and

$$[D_s(e^{vl}-1)]_{v=0} \equiv 0 \pmod{l^2};\quad s>1.$$

Hence $[D_{kl}(Z)]_{v=0}$ is divisible by l^2 and therefore

$$a_k B_{kl} \equiv 0 \pmod{l^2},$$

and using the hypothesis of the lemma we have

$$a_k \equiv 0 \pmod{l},$$

hence by our assumption at the beginning of the proof, $t \not\equiv 0 \pmod{l}$, giving from (3)

$$\eta^t = \delta^l$$

where δ is a unit in $k(\zeta)$. Raise each side to the power t_1, with t_1 a rational integer such that $tt_1 \equiv 1 \pmod{l}$; we have $\eta = \delta_1{}^l$, which proves Lemma 2.

2. We now enunciate

THEOREM I. *Under the following assumptions*:

(1) *the second factor of the class number of the field* $k(\zeta)$ *is prime to* l;

(2) *none of the Bernoulli numbers* B_{nl}, $n = 1, 2, \cdots, (l-3)/2$, *is divisible by* l^3;

the equation (2) *is impossible in case* II.

The proof is, in the main, an extension of Kummer's argument employed in his 1857 memoir already cited here.

We consider the equation

$$\omega^l + \theta^l = \eta\kappa^{ml}\xi^l, \tag{6}$$

where ω, θ and ξ are integers in the field defined by $\zeta + \zeta^{-1}$ and are also prime to each other in that field; η is a unit in this field and

$$\kappa = (1 - \zeta)(1 - \zeta^{-1}).$$

Also the rational integer m is greater than 1. We shall show this equation to be impossible under the assumptions mentioned; it follows easily that (2) is impossible in case II under the same assumptions.

Evidently

$$\prod_{a=0}^{l-1}(\omega + \zeta^a\theta) = \eta\kappa^{ml}\xi^l. \tag{7}$$

We note that from

$$\omega + \zeta^a\theta + (\zeta^r - \zeta^a)\theta = \omega + \zeta^r\theta$$

it is possible to infer that each one of the factors of (7) is divisible by $(1-\zeta)$ but the right hand member of (7) is divisible by a higher power of $\lambda = (1-\zeta)$ than l. Hence one of the factors on the left hand side is divisible by λ^2.

Since $\omega + \theta$ is real it is divisible by κ and since, if $a \not\equiv 0 \pmod{l}$,

$$\omega + \zeta^a\theta = \omega + \theta + (\zeta^a - 1)\theta,$$

we have that the expression on the left is divisible by λ but not by λ^2. Also $\omega+\zeta^a\theta$ and $\omega+\zeta^b\theta$ have no factor in common aside from λ since their difference is $(\zeta^a-\zeta^b)\theta$ and θ is prime to ω, hence

$$\left(\frac{\omega+\zeta^a\theta}{1-\zeta^a}\right) = \mathfrak{j}_a{}^l \qquad (a = 1,2,\cdots,l-1),$$

$$(\omega+\theta) = (\kappa)^{ml-(l-1)/2}\mathfrak{j}_0{}^l.$$

By Lemma I,

$$\left(\frac{\omega+\zeta^a\theta}{1-\zeta^a}\right)\left(\frac{\omega+\zeta^{-a}\theta}{1-\zeta^{-a}}\right)^{l-1} = (\rho_a')^l, \tag{7a}$$

where ρ_a' is an integer in $k(\zeta)$.* But we also have

$$\left(\frac{\omega+\zeta^a\theta}{1-\zeta^a}\right)\left(\frac{\omega+\zeta^{-a}\theta}{1-\zeta^{-a}}\right) = (\mathfrak{j}_a\mathfrak{j}_{-a})^l, \tag{7b}$$

and since $\mathfrak{j}_a\mathfrak{j}_{-a}$ belongs to the field $k(\zeta+\zeta^{-1})$ we have, if h is the class number of $k(\zeta+\zeta^{-1})$,

$$(\mathfrak{j}_a\mathfrak{j}_{-a})^h \sim 1, \tag{7c}$$

with h prime to l; hence if $hh_1 = 1+sl$, s and h_1 rational integers, then

$$(\mathfrak{j}_a\mathfrak{j}_{-a})^{hh_1} \sim (\mathfrak{j}_a\mathfrak{j}_{-a})(\mathfrak{j}_a\mathfrak{j}_{-a})^{sl},$$

which with (7b) and (7c) gives

$$\mathfrak{j}_a\mathfrak{j}_{-a} \sim 1;$$

hence by (7b),

$$\left(\frac{\omega+\zeta^a\theta}{1-\zeta^a}\right)\left(\frac{\omega+\zeta^{-a}\theta}{1-\zeta^{-a}}\right) = (\rho_a'')^l\beta, \tag{7d}$$

where β is a unit and ρ'' an integer in $k(\zeta+\zeta^{-1})$. Multiplying this by (7a) we have

$$\left(\frac{\omega+\zeta^a\theta}{1-\zeta^a}\right)^2\left(\frac{\omega+\theta^{-a}\theta}{1-\zeta^{-a}}\right)^l = \beta(\rho_a''\rho_a')^l$$

and

$$\frac{\omega+\zeta^a\theta}{1-\zeta^a} = \eta_a\rho_a{}^l \qquad (a = 1,2,\cdots,l-1), \tag{8}$$

where η_a is a real unit in $k(\zeta)$, ρ_a an integer in that field. Also since the second factor of the class number of $k(\zeta)$ is prime to l we have

$$\omega+\theta = \eta_0\kappa^{ml-(l-1)/2}\rho_0{}^l, \tag{8a}$$

* Here ρ is not used in the same sense as on p. 617.

where ρ_0 is an integer in $k(\zeta+\zeta^{-1})$. If we write ρ_{-a} for the integer in the field $k(\zeta)$ which is obtained from ζ_a by means of the substitution (ζ/ζ^{-1}), we have from the relation (8) for a and $-a$, and (8a), three relations from which we may eliminate ω and θ and obtain, if η' is a unit in $k(\zeta)$,

$$\rho_a{}^l - \rho_{-a}{}^l = \eta_a'(1-\zeta)^{(2m-1)l}\rho_0{}^l.$$

Decomposing the left hand member into linear factors we note from the above that we have, if $\mathfrak{a}$ is an ideal in $k(\zeta)$

$$\frac{\rho - \zeta^i\rho_{-a}}{1-\zeta^i} = \mathfrak{a}_i{}^l \qquad (i = 1,2,\cdots,l-1); \tag{9}$$

$$\rho_a - \rho_{-a} \equiv 0 \pmod{(1-\zeta)^{(2m-2)l+1}}. \tag{9a}$$

The left hand member of (9) is unaltered by the substitution (ζ/ζ^{-1}) and since the second factor of the class number of $k(\zeta)$ is prime to l we have in the same way that (7d) was obtained

$$\frac{\rho_a - \zeta^i\rho_{-a}}{1-\zeta^i} = \eta_i''\mu_i{}^l \qquad (i = 1,2,\cdots,l-1),$$

where η_i'' is a unit and μ_i is an integer in $k(\zeta+\zeta^{-1})$. This relation with (9a) gives

$$\rho_a \equiv \eta_i''\mu_i{}^l \pmod{\lambda^{(2m-2)l}}.$$

Since $m>1$, then $(2m-2)l \geqq 2l$. Using the relation (8) with the last relation given, we obtain

$$\frac{\omega+\zeta^a\theta}{1-\zeta^a} \equiv \eta_a(\eta_i'')^l\mu^{l^2} \pmod{\lambda^{(2m-2)l}}.$$

We have a similar relation with b in place of a and division of these two relations gives

$$\frac{\eta_a(\eta_i'')^l}{\eta_b(\eta_i^{(3)})^l} \equiv c^{l^2} \pmod{\lambda^{2l}}, \tag{10}$$

where c is a rational integer with η_b and $\eta_i^{(3)}$ units in $k(\zeta)$. By Lemma 2 it follows that the unit on the left hand side is the lth power of a unit in $k(\zeta)$. We may now write, if $\rho_a^* = (\eta_i'')^{-l}\rho_a$,

$$\frac{\omega+\zeta^a\theta}{1-\zeta^a} = \eta_a(\eta_i'')^l\rho_a{}^{*l};$$

from the product of this expression and that obtained after making the substitution (ζ/ζ^{-1}) we have

$$\omega^2 + (\zeta^a + \zeta^{-a})\omega\theta + \theta^2 = \eta_a^*(2 - \zeta^a - \zeta^{-a})\omega^{*l},$$
$$\eta_a^* = (\eta_a(\eta_i')^l)^2$$

and ω^* is an integer in $k(\zeta+\zeta^{-1})$. We have a similar equation for b in place of a and also from (8a)

$$\omega^2 + 2\omega\theta + \theta^2 = \eta_0^2\kappa^{2ml-l+1}\rho_0^{2l}.$$

Elimination of the quantities $\omega^2+\theta^2$ and $\omega\theta$ from these three equations gives, if $a \not\equiv \pm b \pmod{l}$,

$$\eta_a^*\omega^{*l} - \eta_b^*\theta^{*l} = \frac{\eta_0^2\kappa^{2ml-l+1}(\zeta^a + \zeta^{-a} - \zeta^b - \zeta^{-b})\rho_0^{2l}}{(2 - \zeta^a - \zeta^{-a})(2 - \zeta^b - \zeta^{-b})}. \tag{10a}$$

Dividing through by η_a^* we obtain as a coefficient of θ^* the quantity which by (10) we saw was the lth power of a unit in $k(\zeta)$, and noting also that the right hand member may be simplified by using the relation

$$\zeta^a + \zeta^{-a} - \zeta^b - \zeta^{-b} = (\zeta^{-b} - \zeta^{-a})(\zeta^{a+b} - 1),$$

we obtain

$$\omega_1^l + \theta_1^l = \delta\kappa^{(2m-1)l}\xi_1^l, \tag{10b}$$

with δ a unit in $k(\zeta)$. This equation is exactly the same form as (6) with ω_1, θ_1, ξ_1 integers in $k(\zeta)$ prime to each other and $2m-1>1$, since $m>1$. Comparing the value of ξ_1 with the value of ξ we see, however, that since $\xi_1=\rho_0^2$, then ξ_1 necessarily contains a lesser number of distinct prime ideal factors than does ξ, aside from the exceptional case when ξ/ρ_0 is a unit in $k(\zeta)$, in which event

$$\frac{\omega + \zeta^i\theta}{1 - \zeta^i}$$

is a unit in the field $k(\zeta)$ where $i=1, 2, \cdots, l-1$, whence

$$\frac{\omega + \zeta^i\theta}{1 - \zeta^i} = \frac{\zeta^k(\omega + \zeta^{-i}\theta)}{1 - \zeta^{-i}},$$

which gives, using $\omega+\theta\equiv 0 \pmod{l}$,

$$(1 - \zeta^k)(1 - \zeta^i) \equiv 0 \pmod{l},$$

and this is impossible for $l>3$.

Hence we obtain by repetition of the process used in connection with (6a) an unlimited series of integers in the field $k(\zeta)$, ξ_1, ξ_2, ξ_3, $\cdots$, in each of which the number of distinct ideal factors is less than in the preceding, which is impossible. This completes the proof of Theorem I.

3. We now consider

THEOREM II. *Under the assumptions*

(1) *only one of the Bernoulli numbers (say* B_n*) in the set*

$$B_1, B_2, \cdots, B_{l_1} \tag{11}$$

is divisible by l;

(2) *the Bernoulli number* B_{nl} *is not divisible by* l^3;

the equation (2) *is impossible in rational integers none zero.*

For proof suppose first that the second factor of the class number of $k(\zeta)$ is prime to l; it then follows from Theorem I that (2) is impossible in case II in view of Assumption 2 of the present theorem. The theorem follows immediately for case I since it is known* that if (2) is satisfied in this case then at least two of the Bernoulli numbers in the set (11) are divisible by l.

Consider now the only other possibility, that is, when the second factor of the class number is divisible by l in case II. Furtwängler† showed that if f is an arbitrary algebraic field and f' the superfield of f obtained by adjoining ζ, then any class in the irregular class group‡ of f may be represented in the form

$$(c)_f = c_1^{x_1} \cdots c_e^{x_e}$$
$$(x_i = 0, 1, \cdots, l^{h_i} - 1\ ;\ h_1 + \cdots + h_l = h'),$$

where $ql^{h'}$ is the class number of f with q prime to l, and also any class in the irregular class group of f' may be written in the form (if $(c)_f$ is now interpreted in f')

$$(c)_f C,$$

where C is a class in f' such that, if $s_1, \cdots, s_j$ are the substitutions of the relative group of f' with respect to f, then if $C = C_{a_1}$,

$$C_{a_1} C_{a_2} \cdots C_{a_j} = 1,$$

where C_a represents the class of ideals obtained from C by the substitution s on the ideals of C. Let the field f be $k(\zeta+\zeta^{-1})$ and then f' is $k(\zeta)$, so every class in the irregular class group of $k(\zeta)$ may be written in the form

$$c_k C$$

---

* Kummer, Berlin Abhandlungen, Mathematisch-Physikalische Klasse, 1857, pp. 61–65.

† Mathematische Annalen, vol. 63 (1906), pp. 21–22; Pollaczek, Mathematische Zeitschrift, vol. 21 (1924), p. 22.

‡ The irregular class group is obtained by considering the group of classes formed by the qth powers of all the ideals in f. Hence the principal class in the irregular class group consists of ideals which are $l^{h'}$th powers of ideals in f.

where

$$CC_s = 1$$

with $s = (\zeta/\zeta^{-1})$. Now consider from (6)

$$\left(\frac{\omega + \zeta^a\theta}{1 - \zeta^a}\right).$$

This has no prime ideal factor which belongs to the field $k(\zeta+\zeta^{-1})$ since if $\mathfrak{q}$ is such an ideal, then

$$\frac{\theta + \zeta^a\omega}{1 - \zeta^a} \equiv \frac{\theta + \zeta^{-a}\omega}{1 - \zeta^{-a}} \equiv 0 \pmod{\mathfrak{q}}$$

which gives

$$(\zeta^a - \zeta^{-a})\omega \equiv 0 \pmod{\mathfrak{q}},$$

which contradicts the assumption that ω and ξ are prime to each other. Hence all prime ideal factors of $(\omega+\zeta^a\theta)/(1-\zeta^a)$ belong to the C classes and the relative norm of each belongs to the principal class of the irregular class group in $k(\zeta)$. We may then write

$$\left[\left(\frac{\theta + \zeta^a\omega}{1 - \zeta^a}\right)\left(\frac{\theta + \zeta^{-a}\omega}{1 - \zeta^{-a}}\right)\right]^q = \eta_1(\tau')^l$$

and using $qq_1 \equiv 1 \pmod{l}$ since q is prime to l we have, using notation of (8),

$$\frac{\theta + \zeta^a\omega}{1 - \zeta^a}\cdot\frac{\theta + \zeta^{-a}\omega}{1 - \zeta^{-a}} = \eta_a^2\tau_a^l,$$

$$\frac{\theta + \zeta^b\omega}{1 - \zeta^b}\cdot\frac{\theta + \zeta^{-b}\omega}{1 - \zeta^{-b}} = \eta_b^2\tau_b^l,$$

$$\theta^2 + \omega^2 + 2\theta\omega = \eta_0^2\kappa^{2ml-l+1}\tau_0^{2l},$$

with $a \not\equiv \pm b \pmod{l}$, and elimination of the two quantities $\theta^2+\omega^2$ and $\omega\theta$ from these three equations gives

$$\eta_a^2\tau_a^l - \eta_b^2\tau_a^l = \delta'\kappa^{(2m-1)l}(\xi')^l, \tag{12}$$

where δ' is a unit and ξ' an integer in $k(\zeta+\zeta^{-1})$. This relation shows that η_a^2/η_b^2 is a primary unit in $k(\zeta+\zeta^{-1})$. But we shall now show that if the second factor of the class number is divisible by l this is contrary to the second assumption of our theorem.

Using the assumption as to the second factor it follows that E_n is the lth power of a unit in $k(\zeta)$, but that $E_{n'}, n' \not\equiv n$, is not an lth power. Suppose

that there exists a primary unit η which is not the lth power of a unit in $k(\zeta)$, we may then set

$$\eta^t = E_1^{s_1}E_2^{s_2}\cdots E_{l_1}^{s_{l_1}}.$$

As in the proof of Lemma II, we may replace this by an identity in e^v of the type (3b). Differentiating $2k$ times and setting $v=0$ we have, using (4a) and (5),

$$s_iB_i \equiv 0 \pmod{l},$$

and $s_i \equiv 0 \pmod{l}$ except for $i=n$. Now $E_n=(E')^l$. Hence η^t is the lth power of a unit with η not the lth power of a unit, hence $t\equiv 0 \pmod{l}$, and as in the proof of Lemma II, we have $s_n \not\equiv 0 \pmod{l}$. It then follows that $\eta^t \equiv c^t \pmod{l^2}$ and we find as from (3b), $B_{nl}\equiv 0 \pmod{l^3}$ which contradicts Assumption 2 of our theorem. Hence in (12), η_a^2/η_b^2 is the lth power of a unit in $k(\zeta)$ and (12) reduces to the same form as (6). Repetition of this process leads to a contradiction as in the proof of Theorem I. This completes the proof of Theorem II.

4. We now proceed to

THEOREM III. *If $l\equiv 1 \pmod 4$ and all the numbers in* (11) *which are divisible by l have even subscripts, then* (2) *is impossible in rational integers none zero.*

This proof depends, in part, on a remarkable device employed by Mirimanoff.* Consider first the relation (2) in case I. Then Kummer, as pointed out in the proof of Theorem II, proved that, if (2) holds in this case, $B_{l_1}\equiv B_{l_1-1} \equiv 0 \pmod{l}$ and one of these subscripts is odd. This proves the theorem for case I.

In case II, consider the generalized form (6) again. Using the result employed in the proof of Theorem II, that all ideal divisors of $(\omega+\zeta^a\theta)/(1-\zeta^a)$ are such that their relative norms with respect to the field $k(\zeta+\zeta^{-1})$ belong to the principal class of the irregular class group, we obtain as before

$$\left(\frac{\theta+\zeta^a\omega}{1-\zeta^a}\right)\left(\frac{\theta+\zeta^{-a}\omega}{1-\zeta^{-a}}\right) = \eta_a\tau_a{}^l,$$

and this may be written, if $\kappa_a=(1-\zeta^a)(1-\zeta^{-a})$,

$$(\theta+\omega)^2 - (\kappa_a)\theta\omega = \eta_a\kappa_a\tau_a{}^l, \tag{13}$$

$$(\theta+\omega)^2 = \eta_0\kappa^{2ml-l+1}\tau_0{}^l. \tag{13a}$$

Now, following Mirimanoff, we may assume, since $l\equiv 1 \pmod 4$, that $(\theta+\omega)^2$

* Crelle, vol. 111 (1893), pp. 26–30.

and $\theta\omega$ belong to the sub-field of $k(\zeta)$ defined by $(\theta+\theta^{-1}+\theta^{r^w}+\theta^{-r^w})$ which we shall call k'; $w=(l-1)/4$. Suppose that the distinct Bernoulli numbers in the set (11) which are divisible by l are

$$B_{a_1}, B_{a_2}, \cdots, B_{a_s}.$$

By the hypothesis of the theorem the a's are all even. Now according to Pollaczek* there exists a fundamental system of l_1 real units in $k(\zeta)$ such that

$$\beta_i^{s-r^{2i}} = (\beta_i')^l \qquad (i = 1,2, \cdots, l_1), \tag{13b}$$

s standing for the substitution (ζ/ζ^r), and r is a primitive root of l. It was also shown by the writer† that there exists a certain system of independent real units in $k(\zeta)$,

$$\alpha_1, \alpha_2, \cdots, \alpha_{l_1},$$

which have the property

$$\alpha_i^{s-r^{2i}} = (\alpha_i')^l,$$

and also

$$\alpha_i = \beta_i^{c_i}(\beta_i'')^l,$$

c_i a rational integer and β_i'' a unit in $k(\zeta)$. As the E's as defined in the present paper also have this property, it follows as in the Annals paper that

$$E_i = \beta_i^{d_i}(\beta_i^{(3)})^l, \tag{14}$$

where $\beta_i^{(3)}$ is a unit in $k(\zeta)$ and d_i is a rational integer.

Now consider the primary units in $k(\zeta)$. Any such may be written in the form

$$s + \sigma l = \gamma = \prod_{i=1}^{l_1} \beta_i^{g_i}, \tag{15}$$

where s is a rational integer and σ an integer in $k(\zeta)$. As in the proof of Lemma II, we may replace this by an identity in e^v and obtain

$$(s + l\sigma(e^v))^{l-1} = \prod_{i=1}^{l_1} \gamma_i^{g_i}(e^v) + \frac{e^{vl}-1}{e^v-1}X \tag{15a}$$

where

$$\gamma_i \equiv 1 \pmod{\lambda},$$

and X is a polynomial in e^v with rational integral coefficients. Now using (14), replacing it by an identity in e^v, differentiating $2i$ times and setting $v=0$, we have from (4a)

* Loc. cit., p. 7.

† Annals of Mathematics, 1929.

$$(16) \qquad (-1)^{i-1}\frac{B_i(r^{2i}-1)}{4i}(l-1) \equiv d_i\left[\frac{d^{2i}\log \beta_i(e^v)}{dv^{2i}}\right]_{v=0},$$

modulo l. We have similarly, $t<l-1$,

$$(17) \qquad (l-1)\left[\frac{d^t \log \beta_i(e^v)}{dv^t}\right]_{v=0} \equiv \frac{d^t \log \gamma_i(c^v)}{dv^t} \pmod{l}.$$

We also have from (13b) after replacing it by an identity in e^v

$$(18) \qquad \left[\frac{d^t \log \beta_i(e^v)}{dv^t}\right]_{v=0} \equiv 0 \pmod{l}$$

for $t \neq 2i$. Differentiating (15a) $2i$ times and setting $v=0$, we have, using (17) and (18),

$$g_i\left[\frac{d^{2i}\beta_i(e^v)}{dv^{2i}}\right]_{v=0} \equiv 0 \pmod{l},$$

and employing (16) we see that $d_i \not\equiv 0 \pmod{l}$; $g_i \equiv 0 \pmod{l}$ except for $i=a_1, a_2, \cdots, a_e$. Hence, from (15), any primary unit in $k(\zeta)$ may be written in the form

$$\gamma = \prod_\alpha \beta_\alpha^{g_\alpha} \delta^l$$

where δ is a unit in $k(\zeta)$ and α ranges over the set $a_1, \cdots, a_e$.

Now consider the unit η_a which occurs in (13). We may write it in the form

$$\eta_a = \prod_{i=1}^{l_1} \beta_i^{h_i'}.$$

Let $h_i' \equiv 2h_i \pmod{l}$, where h_i is a rational integer, and we have

$$\eta_a = \prod_{i=1}^{l_1} \beta_i^{2h_i}\delta_1^l, \quad 0 \leqq h_i < l,$$

where δ_1 is a unit in $k(\zeta)$. The relation (13) then becomes

$$(19) \qquad (\theta+\omega)^2 - (\kappa_a)\theta\omega = \kappa_a \prod_i \beta_i^{2h_i}(\phi_a')^l.$$

Set

$$\theta = \prod_{i=1}^{l_1} \beta_i^{h_i}\beta_i^{h_i}(\zeta^{r^w})\theta', \quad \omega = (\theta/\theta')\omega',$$

where $\beta_i(\zeta^{r^w})$ represents the unit obtained from β_i by the substitution (ζ/ζ^{r^w}) with $w=(l-1)/4$. Now $\beta_i\beta(\zeta^{r^w})$ belongs to the field k'. Hence

$(\theta'+\omega')^2$ and $\theta'\omega'$ belong to this field, since $(\theta+\omega)^2$ and $\theta\omega$ do. From (19) we then have

$$(\theta' + \omega')^2 - (\kappa_a)\theta'\omega' = \kappa_a \frac{\prod_i \beta_i^{h_i}}{\prod_i \beta_i^{h_i}(\zeta^{r^w})} (\phi_a')^l. \tag{20}$$

We have

$$\beta_i(\zeta^r) = \beta_i r^{2i}\delta_2{}^l,$$

where δ_2 is a unit in $k(\zeta)$. Hence

$$\beta_i(\zeta^{r^w}) = \beta_i r^{2iw}\delta_3{}^l = \beta_i\delta_4{}^l$$

if i is even and the δ's are units in $k(\zeta)$. If i is odd we have

$$\beta_i(\zeta^{r^w}) = \beta_i r^{i(l-1)/2}(\delta_5')^l.$$

Hence (20) becomes

$$(\theta' + \omega')^2 - (\kappa_a)\theta'\omega' = \kappa_a \prod_s \beta_s{}^{h_s(1-r^{s(l-1)/2})}\phi_a{}^l, \tag{21}$$

where s ranges over all the odd integers in the set $1, 2, \cdots, l_1$. Dividing through by κ_a, we have, using $\theta'+\omega' \equiv 0 \pmod{l^2}$,

$$\theta'\omega' \equiv \prod_s \beta_s{}^{h_s(1-r^{s(l-1)/2})}\phi_a{}^l \pmod{l}.$$

Setting ζ^{r^w} for ζ in this relation we have, since $\theta'\omega'$ is unaltered by this substitution,

$$\prod_s \beta^{2h_s(r^{s(l-1)/2}-1)} \equiv c \pmod{l},$$

where c is a rational integer. Since all the Bernoulli numbers divisible by l have even subscripts and none can be an s and all the g's in (15) are divisible by l except for $i=a_1, a_2, \cdots, a_e$, then

$$2h_s(r^{s(l-1)/2} - 1) \equiv 0 \pmod{l}$$

and $h_s=0$, so (21) becomes

$$(\theta' + \omega')^2 - (\kappa_a)\theta'\omega' = \kappa_a\phi_a{}^l.$$

Setting ζ^{r^w} for ζ in this relation gives

$$(\theta' + \omega')^2 - (2 - \zeta^{ar^w} - \zeta^{-ar^w})\theta'\omega' = \kappa_a(\zeta^{r^w})(\phi_a(\zeta^{r^w}))^l.$$

Taking these two equations with (13a) and eliminating $(\theta'+\omega')^2$ and $\theta'\omega'$ we obtain

$$\phi_a{}^l - (\phi_a(\zeta^{r^w}))^l = \gamma\kappa^{(2m-1)l}\tau_0{}^l,$$

where γ is a unit in $k(\zeta)$. This equation belongs to the same class as (6) and we obtain an equation therefrom of the same form as (13) and having the property that $(\phi_a-\phi_a(\zeta^{r^w}))^2$ and $\phi_a\phi_a(\zeta^{r^w})$ each belong to the field k'. We may then proceed as in the last part of the proof of Theorem I and ultimately reach a contradiction.

5. We may now proceed to

THEOREM IV. *Under the following assumption*:

None of the units E_a, $k=a_1, a_2, \cdots, a_s$, is congruent to the lth power of an integer in the field $k(\zeta)$ mod $\mathfrak{p}$, *where $\mathfrak{p}$ is a prime ideal divisor of p, p is a prime $<(l^2-l)$ of the form* 1 mod l, *and $a_1, a_2, \cdots, a_s$ are the subscripts in the Bernoulli numbers in the set* (11) *which are divisible by l;*
the relation (2) *is impossible in case* II.

To prove this we note first that it follows from this assumption that the second factor* of the class number of $k(\zeta)$ is prime to l. Instead of the generalized equation (6a) we shall here consider the same form with $\eta=1$ and ω, θ and ξ integers in the field $k(\zeta+\zeta^{-1})$ prime to each other, that is, we consider the equation

$$\omega^l + (\kappa^m\xi)^l = -\theta^l.$$

Set

$$-\kappa^m\xi = \phi,$$

so that

$$\omega^l + \phi^l = -\theta^l. \tag{22}$$

As in the proof of Theorem I we note that $(\omega+\zeta^a\phi)(\omega+\zeta^a\phi)^{l-1}$ is a primary number in $k(\zeta)$, so that we have as in the said proof

$$\omega + \zeta^a\phi = \eta_a\sigma_a{}^l \qquad (a = 1,2, \cdots, l-1), \tag{23}$$

where η_a is a real unit in $k(\zeta)$. Hence we have

$$\sigma_{-a}{}^l(\omega + \zeta^a\phi) = \sigma_a{}^l(\omega + \zeta^{-a}\phi). \tag{23a}$$

Now θ is prime to $\mathfrak{p}$, for if divisible by $\mathfrak{p}$, we may assume $\mathfrak{p}$ prime to ω and we obtain from (22), in the same manner in which (23) was derived,

$$\theta + \zeta^a\phi = \gamma_a\tau^l, \tag{24}$$

where γ_a is a real unit and τ an integer in $k(\zeta)$. Since θ is divisible by $\mathfrak{p}$ it is also divisible by $\mathfrak{p}_{-1}$ since θ belongs to the field $k(\zeta+\zeta^{-1})$. Hence using power characters† we have

* Stafford and Vandiver, Proceedings of the National Academy of Sciences, 1930; Vandiver, Bulletin of the American Mathematical Society, vol. 35 (1929), pp. 333–35.

† Hilbert, loc. cit., p. 365.

$$\left\{\frac{\zeta^a\phi}{\mathfrak{p}\mathfrak{p}_{-1}}\right\} = \left\{\frac{\gamma_a}{\mathfrak{p}\mathfrak{p}_{-1}}\right\},$$

and since ϕ and γ_a are in the field $k(\zeta+\zeta^{-1})$ we obtain

$$\left\{\frac{\zeta^a}{\mathfrak{p}\mathfrak{p}_{-1}}\right\} = \left\{\frac{\zeta}{\mathfrak{p}}\right\}^{2a} = 1,$$

and $p-1\equiv 0 \pmod{l^2}$, which contradicts our hypothesis concerning p. Hence the σ's are prime to $\mathfrak{p}$. Hence we obtain from (23) by raising both members to the power c, $p=1+cl$,

$$(\omega + \zeta^a\phi)^c \equiv (\omega + \zeta^{-a}\phi)^c \pmod{\mathfrak{p}}, \tag{24a}$$

or

$$\omega^c + c\omega^{c-1}\zeta^a\phi + \cdots + \zeta^{ac}\phi^c \equiv \omega^c + c\omega^{c-1}\zeta^{-a}\phi + \cdots + \zeta^{-ca}\phi^c \pmod{\mathfrak{p}}.$$

Multiply both sides by ζ^a. We have, after setting $a=0$ (for which the above relation obviously holds), 1, 2, $\cdots$, $l-1$, and adding the resulting congruences,

$$\phi\omega^{c-1} \equiv 0 \pmod{\mathfrak{p}} \tag{25}$$

under the assumption that $c<l-1$. Now ω is not divisible by $\mathfrak{p}$, as we can obtain a contradiction by assuming it divisible by $\mathfrak{p}$ and employing (23) in the same way that (24) was used. Hence we obtain from (25)

$$\phi \equiv 0 \pmod{\mathfrak{p}}. \tag{25a}$$

Since $p=1+cl$, it is of the first degree and we may therefore write

$$p = \mathfrak{p}_1\mathfrak{p}_2 \cdots \mathfrak{p}_{l-1},$$

the $\mathfrak{p}$'s being distinct ideals. Now $\mathfrak{p}$ was an arbitrary prime ideal divisor of p and since p is not divisible by the square of an ideal it follows that we may write in lieu of (25a)

$$\phi \equiv 0 \pmod{p}.$$

We shall now show that $(\omega+\theta)$ is divisible by (p). In order to effect this we again consider (22) and obtain, exactly as in the proof of Theorem I,

$$\frac{\omega + \zeta^a\theta}{1 - \zeta^a} = \eta_a\rho_a{}^l, \tag{26}$$

$$\omega + \theta = \eta_0\kappa^{ml-(l-1)/2}\rho_0{}^l, \tag{26a}$$

where the η's are units and the ρ's integers in $k(\zeta)$. Since ϕ is divisible by p let us assume that

$$\omega + \zeta^k\theta \equiv 0 \pmod{\mathfrak{p}_n},$$

where $\mathfrak{p}_n$ is a prime ideal divisor of p for

$$k \not\equiv 0 \pmod{l}.$$

Make the substitution in this congruence $(\zeta/\zeta^{k'})$ where

$$kk' \equiv 1 \pmod{l}.$$

By this substitution $\mathfrak{p}_n$ is changed into another prime ideal divisor of p which we shall call $\mathfrak{p}$, hence, if ω' and θ' are the corresponding conjugates of ω and θ,

$$\omega' + \zeta\theta' \equiv 0 \pmod{\mathfrak{p}}. \tag{27}$$

It then follows that $\omega'+\zeta^a\theta'$ is prime to $\mathfrak{p}$ for $a \not\equiv 1$, and, from (26),

$$\frac{\omega'+\zeta^k\theta'}{\zeta^k - 1}\rho_{-k}^{\prime l} = \frac{\omega'+\zeta^{-k}\theta'}{\zeta^{-k} - 1}\rho_k^{\prime l}.$$

If we assume now that k is any integer in the set $1, 2, \cdots, l-1$ excepting unity, we may raise both sides of this relation to the power c and obtain, since the ρ's are prime to p,

$$(\omega' + \zeta^k\theta')^c \equiv (\zeta^k\omega' + \theta')^c \pmod{\mathfrak{p}}, \tag{27a}$$

$$k = 2, 3, \cdots, l-1.$$

We also have

$$\omega' + \zeta^k\theta' - \omega' - \zeta\theta' \equiv \theta'(\zeta^k - \zeta) \pmod{\mathfrak{p}},$$

and from (27) and (27a),

$$\left\{\frac{\theta'(\zeta^k - \zeta)}{\mathfrak{p}}\right\} = \left\{\frac{\theta'(1 - \zeta^{k+1})}{\mathfrak{p}}\right\},$$

the symbol denoting the lth power character, or

$$\left\{\frac{\zeta}{\mathfrak{p}}\right\}\left\{\frac{\zeta^{k-1} - 1}{\mathfrak{p}}\right\} = \left\{\frac{\zeta^{k+1} - 1}{\mathfrak{p}}\right\},$$

whence

$$\left\{\frac{\zeta}{\mathfrak{p}}\right\}\left\{\frac{\frac{\zeta^{k-1} - 1}{\zeta - 1}}{\mathfrak{p}}\right\} = \left\{\frac{\frac{\zeta^{k+1} - 1}{\zeta - 1}}{\mathfrak{p}}\right\}. \tag{28}$$

Now consider the expression

$$\left\{\frac{\zeta^t - 1}{\zeta - 1}\right\} = \zeta^T,$$

where

$$T = \operatorname{ind}\left(\frac{\zeta^t - 1}{\zeta - 1}\right).$$

Using a known formula* we have

$$\operatorname{ind}\left(\frac{\zeta^t - 1}{\zeta - 1}\right) \equiv \frac{1}{2}(t-1)\operatorname{ind}\zeta - 2\sum_{n=1}^{l_1}\frac{\operatorname{ind} E_n(\zeta^t) - \operatorname{ind} E_n(\zeta)}{r^{2n} - 1} \pmod{l}. \tag{29}$$

Now if

$$t \equiv r^i \pmod{l},$$

then because of the relation†

$$E_n{}^{s^i - r^{2in}} = \eta^l\ ;\ \eta \text{ a unit in } k(\zeta),$$

we may write

$$\operatorname{ind} E_n(\zeta^t) \equiv t^{2n} \operatorname{ind} E_n(\zeta) \pmod{l}. \tag{29a}$$

Substituting in (29) we obtain

$$\operatorname{ind}\left(\frac{\zeta^t - 1}{\zeta - 1}\right) \equiv \frac{1}{2}(t-1)\operatorname{ind}\zeta - 2\sum_{n=1}^{l_1}\frac{(t^{2n} - 1)\operatorname{ind} E_n(\zeta)}{r^{2n} - 1} \pmod{l}. \tag{29b}$$

Now apply this relation to (28). We obtain

$$\sum_{n=1}^{l_1}\frac{((k-1)^{2n} - 1)\operatorname{ind} E_n(\zeta)}{r^{2n} - 1} \equiv \sum_{n=1}^{l_1}\frac{((k+1)^{2n} - 1)\operatorname{ind} E_n(\zeta)}{r^{2n} - 1} \pmod{l}.$$

Now expand the powers of $k-1$ and $k+1$ by the binomial theorem and arrange the results in descending powers of k. We have, after noting that the coefficients of the even powers of k are the same on both sides of the congruence,

$$A_{l-4}k^{l-4} + A_{l-6}k^{l-6} + \cdots + A_1 k \equiv 0 \pmod{l}, \tag{30}$$

where the A's are expressions involving t, r, and $\operatorname{ind} E_n(\zeta)$. We have

$$A_{l-4} = \frac{2(l-3)}{r^{l-3} - 1}\operatorname{ind} E_{l_1}(\zeta).$$

Formula (30) is true for all values of k excepting $k=1$. If we put $k=2$, 4,

* Kummer, Crelle's Journal, vol. 56, p. 277.

† Here we are using the Kronecker-Hilbert notation of symbolic powers, $E_n{}^{s^i}$ corresponding to the substitution $E_n(\zeta)/E_n(\zeta^{r^i})$.

$\cdots, l-3$, in turn, we obtain l_1 congruences from which we infer that (cf. (31a))

$$0 \equiv A_1 \equiv A_3 \equiv A_5 \cdots \equiv A_{l-4} \pmod{l},$$

since the determinant formed by the coefficients of the A's is prime to l. From the coefficient of A_{l-4} already found we obtain

$$\operatorname{ind} E^{l_1}(\zeta) \equiv 0 \pmod{l}.$$

Also

$$A_{l-6} = 2\binom{l-3}{3}\frac{\operatorname{ind} E_{l_1}(\zeta)}{r^{l-3}-1} + \frac{2(l-5)\operatorname{ind} E_{l_1-1}(\zeta)}{r^{l-5}-1}.$$

From this we obtain

$$\operatorname{ind} E_{l_1-1}(\zeta) \equiv 0 \pmod{l}.$$

Proceeding in this way we find that all the E's are such that

$$\operatorname{ind} E_n \equiv 0 \pmod{l},$$

which contradicts the first assumption in our theorem. From this it easily follows that

$$\left\{\frac{E_n}{\mathfrak{p}_i}\right\} = 1 \qquad (i = 1,2, \cdots, l-1);$$

using (29a), hence, $\omega+\zeta^k\theta \not\equiv 0 \pmod{\mathfrak{p}_i}$; $k \not\equiv 0 \pmod{l}$. We have then completely established the fact that

$$\omega + \theta \equiv 0 \pmod{p}$$

in (26a). Because of this relation we obtain from (26a)

$$\rho_0 \equiv 0 \pmod{p}.$$

Taking (26) and multiplying it by the equation which was obtained from it by a substitution (ζ/ζ^{-1}), then substituting b in lieu of a in the result and also squaring equation (26a), we obtain three equations from which we may eliminate $\omega^2+\theta^2$ and $\omega\theta$ as was done in the proof of Theorem I. We obtain an equation of the type (10a), the only possible difference being in the unit which appears in the right hand member. If we transpose the term $\eta_b^*\theta^{*l}$ to the right hand member and take lth power characters of each member we obtain, since

$$\rho_0 \equiv 0 \pmod{p},$$

$$\left\{\frac{\eta_a^*}{\mathfrak{p}}\right\} = \left\{\frac{\eta_b^*}{\mathfrak{p}}\right\},$$

where $\mathfrak{p}$ is a prime ideal divisor of p. From this we have

$$\left\{\frac{\eta_a^* \eta_b^{*l-1}}{\mathfrak{p}}\right\} = 1.$$

In transforming this expression we shall employ a method which has proved of great use in handling a number of questions in cyclotomic field theory. Since the E's form a system of independent units in $k(\zeta)$ we may set

$$(\eta_a^* \eta_b^{*l-1})^d = E_1^{d_1} E_2^{d_2} \cdots E_{(l-3)/2}^{d_{(l-3)/2}}. \tag{31}$$

Suppose now that d is divisible by l; then we may assume, as at the beginning of the proof of Lemma II, that not all the d's are divisible by l. Hence by a previous paper of the writer's† we infer that $E_s^{d_s}$ is the lth power of a unit in $k(\zeta)$, $s=1, 2, \cdots, l_1$, where $l_1=(l-3)/2$, and therefore $E_s=\gamma^l$, where γ is a unit in $k(\zeta)$, for some s. Then, as we have already seen during the course of the proof of Theorem III, s is included in the set $a_1, a_2, \cdots, a_e$. But this gives

$$\left\{\frac{E_s}{\mathfrak{p}}\right\} = 1,$$

contrary to the assumption stated in Theorem IV. Hence

$$d \not\equiv 0 \pmod{l}.$$

Then in the revised equation (10a) we can also set ζ^k in lieu of ζ; $k=1, 2, \cdots, l-1$. In view of the relation (29a) we have, from (31),

$$d \operatorname{ind}(\eta_a^*(\zeta^k)\eta_b^*(\zeta^k)^{l-1}) \equiv \sum_{n=1}^{l_1} d_n k^{2n} \operatorname{ind} E_n \equiv 0 \pmod{l}, \quad k = 1,2, \cdots, l-1.$$

We obtain l_1 congruences, and since the determinant

$$\begin{vmatrix} 1 & 1 & 1 & 1 \\ 2^2 & 2^4 & 2^6 \cdots & 2^{l-3} \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ l_1^2 & l_1^4 & \cdots & l_1^{l-3} \end{vmatrix} = \prod_{i,j}^{1 \text{ to } l_1} (i+j)(i-j)(l_1!)^2, \quad i>j, \tag{31a}$$

is prime to l we obtain

$$d_n \operatorname{ind} E_n \equiv 0 \pmod{l}; \quad n = 1,2, \cdots, l_1.$$

† Bulletin of the American Mathematical Society, vol. 35 (1929), pp. 333–335.

Now also the unit $\eta_a^*\eta_b^{l-1}$ is a primary unit in $k(\zeta)$. This follows immediately from the revised equation (10a). Hence the only E's which can appear in (31) are those referred to in the Assumption of our Theorem IV which shows that all the d's appearing as exponents are divisible by l. Hence $(\eta_a^*\eta_b^{*l-1})^d$ and therefore $\eta_a^*\eta_b^{*l-1}$ is the lth power of a unit in $k(\zeta)$. Now using our revised equation (10a) and dividing through by η_a^*, we obtain

$$\omega_1^l + \theta_1^l = \delta_1\kappa^{(2m-1)l}\xi_1^l, \tag{32}$$

where ω_1, θ_1, ξ_1 are integers in $k(\zeta)$ and δ_1 is a unit in that field. Also the first three integers mentioned are prime to each other in $k(\zeta)$ with ξ_1 divisible by p. This equation is a more generalized form than (22). However we may proceed as in the treatment of (22), obtaining

$$\frac{\omega_1 + \zeta^a\theta_1}{1-\zeta^a} = \alpha_a\nu_a^l,$$

$$\omega_1 + \theta_1 = \alpha_0\kappa^{(2m-1)l-(l-1)/2}\nu_0^l,$$

where $a = 1, 2, \cdots, l-1$; the α's are units and the ν's integers in $k(\zeta)$. Since ξ_1 is divisible by p it follows that either ν_0 or ν_a is divisible by a prime ideal divisor of p and we may show exactly as in our treatment of (26) and (26a) that $\omega_1+\theta_1$ is divisible by p. Also eliminating ω_1 and θ_1 as ω and θ were eliminated from the equations based on (26) and (26a) and showing that the units involving α are lth powers of units in $k(\zeta)$ as it was proved that the units involving the η's were lth powers, we find

$$\omega_2^l + \theta_2^l = \delta_2\kappa^{(4m-3)l}\xi_2^l, \tag{32a}$$

with ξ_2 divisible by p and prime to the integers ω_2 and θ_2, and δ_2 being a unit in the field. This equation is of precisely the same form as (32), hence by repetition of this process we obtain as in the proof of Theorem I an unlimited series of ideals in each of which the number of distinct prime ideal factors is less than in the preceding, which is impossible. This completes the proof of Theorem IV.

6. We shall now consider the equation (2) with x, y and z rational integers, and derive criteria for the solution of it in such a way that the fact that x, y and z are rational integers is employed during the argument. As noted in the introduction to this paper this is apparently the first time criteria of this type have been obtained for the second case of Fermat's last theorem. The theorem is as follows:

THEOREM V. *Under the following assumptions*:
(1) *there exists a rational prime integer* p *such that the congruence*

$$u^l + v^l + w^l \equiv 0 \pmod{p}$$

has no solution u, v *and* w *all rational integers prime to* p, *and* $p \not\equiv 1 \pmod{l^2}$;
(2) *the relation*

$$\left\{\frac{E_a}{\mathfrak{p}}\right\} \neq 1$$

holds, where a *ranges over the values* $a_1, a_2, \cdots, a_s$, *these integers being the subscripts of Bernoulli numbers in the set* (11) *which are divisible by* l, *and* $\mathfrak{p}$ *is a prime ideal divisor of* p;
the equation (2) *is impossible in rational integers none zero.*

For proof consider the equation (2). If this holds then by the first assumption in our theorem it follows that one of the integers x, y, and z is divisible by p. If none of these integers is divisible by l then the theorem follows because of Theorem IV in a previous paper.* If we suppose that one of these integers, say z, is divisible by l, then it follows that z is divisible by p since if x or y is divisible by p we have by Furtwängler's Theorem

$$p^{l-1} \equiv 1 \pmod{l^2},$$

which contradicts the assumption that

$$p \not\equiv 1 \pmod{l^2},$$

for it is known that in order that Assumption 1 of the Theorem hold we must have

$$p \equiv 1 \pmod{l}.$$

We now proceed with equation (2) in the same manner as in the treatment of (32), and we find that $x+y$ is divisible by p. We then find as before an equation of the type (32a) with ξ_2 divisible by p and we ultimately reach a contradiction as described previously.

7. We now consider the application of these theorems to Fermat's last theorem for special exponents. The work of Kummer concerning the examination of special exponents has been described in the introduction to this paper. It was also stated in the introduction that the primes l less than 211 were tested as to being regular. The main formula employed for this is

$$\frac{1 - 3^{l-2a} - 4^{l-2a} + 6^{l-2a}}{4a}(-1)^a B_a \equiv \sum_{s=[l/6]+1}^{[l/4]} s^{2a-1} \pmod{l}.$$

* Annals of Mathematics, (2), vol. 27 (1926), p. 55.

The proof of this formula and the details of the computations will be given in a paper by Professor E. T. Stafford and myself.* As noted before, the only irregular exponents found within the range mentioned were those which had previously been discovered by Kummer. Hence we may state

THEOREM VI. *The equation*

$$\alpha^l + \beta^l + \gamma^l = 0$$

is impossible for all odd prime exponents l less than 211 excepting possibly $l=37, 59, 67, 101, 103, 131, 149$ and 157 if α, β and γ are integers in the field $k(\zeta)$ none zero.

As already noted only one of the Bernoulli numbers in the set (11) is divisible by l in each of the cases $l=37, 59, 67, 101, 103, 131, 149$. We then apply Theorem II to these cases by employing the formula given by

LEMMA 3. *If $1<a<(l-1)/2$ and l is a prime integer >5 then*

$$\frac{(-1)^{a-1}B_{al}(2^{2al}-1)}{2al} \equiv 1^{2al-1} + 3^{2al-1} + 5^{2al-1} + \cdots + (l-2)^{2al-1} \equiv B'_{al} \pmod{l^2} \tag{33}$$

and

$$\frac{(-1)^{a}B_{al}(2^{2al}-1)}{2^{2al}al} = A_a \equiv 1^{2al-1} + 2^{2al-1} + \cdots + \left(\frac{l-1}{2}\right)^{2al-1} \pmod{l^2}. \tag{33a}$$

To prove these relations we employ the Bernoulli summation formula

$$s_n(l) = lb_n + l\frac{n}{2}(lb_{n-1}) + l^2\frac{n(n-1)}{2\cdot 3}(lb_{n-2}) + \cdots + l\frac{l^{r-1}}{r+1}\binom{n}{r}(lb_{n-r}) + \cdots,$$

where

$$s_n(l) = 1^n + 2^n + \cdots + (l-1)^n,$$

and

$$b_{2a} = (-1)^{a-1}B_a, \quad b_{2a+1} = 0, \quad a > 0,$$

and n is an arbitrary positive integer.

Set $n=2al$. The second term in the right hand member of the summation formula is zero. The third term is divisible by l^4, since by the von Staudt-Clausen Theorem lb_{n-2} is either a fraction divisible by l or, if $n-2$ is divisible by $l-1$, then l appears to the first power in the denominator of b_{n-2}, but the

* Proceedings of the National Academy of Sciences, 1930.

latter condition can not hold for $n=2al$. Also it is easily seen by induction that $l^{r-1}>l^2\,(r+1)$ for $l>5$, and $r>3$. Hence $l^{r-1}/(r+1)$ is divisible by l^3 and the corresponding term in the above expansion is divisible by l^4, hence $s_n(l)\equiv lb_n \pmod{l^4}$ for $n=2al$.

We now introduce the formula*

$$\frac{(n^i-1)\sum_{a=1}^{p-1}a^i}{p}=\sum_{a=1}^{p-1}\sum_{s=1}^{i}\binom{i}{s}a^i\left(\frac{y_a}{a}\right)^s p^{s-1},$$

where i is an arbitrary integer, n is an integer prime to p and not unity, and

$$y_a\equiv -\frac{a}{p} \pmod{n},\quad 0\leqq y_a<n.$$

Setting $n=2$ in this formula, $i=2al$ and $p=l$, and employing the congruence just obtained, we have

$$(-1)^{a-1}(2^{2al}-1)B_{al}\equiv(1^{2al-1}+3^{2al-1}+\cdots+(l-2)^{2al-1})2al$$
$$+(1^{2al-2}+3^{2al-2}+\cdots+(l-2)^{2al-2})\frac{2al(2al-1)}{2}l \pmod{l^3}.$$

It is known that

$$1^k+3^k+\cdots+(l-2)^k\equiv 0 \pmod{l}$$

if k is even. Consequently the last relation reduces to (33). Also it is known that

$$1^n+2^n+\cdots+(l-1)^n\equiv 0 \pmod{l^2}$$

if n is odd and $n\not\equiv 1 \pmod{(l-1)}$. Hence subtracting this congruence from the preceding we obtain (33a), since $2al-1\not\equiv 1 \pmod{(l-1)}$. The congruence (33a) was previously obtained for the case n arbitrary by Mirimanoff† who gave a different proof.

For $l=37$ the right-hand member of (33), with $a=16$, was computed directly by Mrs. A. C. S. Williams, who found

$$B'_{16\cdot 37}\equiv 42\cdot 37 \pmod{37^2}.$$

This agrees with Kummer's‡ calculations, as he found

$$\frac{B_{16\cdot 37}}{37}\equiv 35\cdot 37 \pmod{37^2}.$$

* Vandiver, Annals of Mathematics, (2), vol. 18 (1925), p. 112, relation (7a).

† Crelle's Journal, vol. 115 (1895), p. 300.

‡ 1857 Article, p. 73.

Similarly Mrs. Williams found that

$$B'_{22\cdot 59} \equiv 59\cdot 17 \pmod{59^2},$$

which also checks with Kummer's computations for that prime. For the case $l=67$ she found $B'_{9\cdot 67}\equiv 67\cdot 13 \pmod{67^2}$. This gives $B_{29\cdot 67}\equiv 67^2\cdot 41 \pmod{67^2}$, in lieu of Kummer's result that $B_{29\cdot 67}\equiv 67^2\cdot 49 \pmod{67^3}$. In my opinion this indicates a misprint in Kummer's article. Also, the computations of Mrs. Williams concerning these three primes were all checked by the writer.

An excellent check on the computations of the A's and B''s is the fact that we must find them in each case divisible by l.

For the known irregular primes >100 and <211 the A's in (33a) were computed by Mr. S. S. Wilks using Monroe and Marchant electrical computing machines. He found $A_{34}\equiv 96\cdot 101 \pmod{101^2}$; $A_{12}\equiv 10\cdot 103 \pmod{103^2}$; $A_{11}\equiv 103\cdot 131 \pmod{131^2}$; $A_{65}\equiv 133\cdot 149 \pmod{149^2}$. In the case 157 where $B_{31}\equiv B_{55}\equiv 0 \pmod{l}$ Mr. Wilks found $A_{31}\equiv 39\cdot 157 \pmod{157^2}$ and $A_{55}\equiv 156\cdot 157 \pmod{157^2}$.

All the computations concerning the B''s and A's were abbreviated by a method which I shall describe for the case 67. We note that the right hand member of (33) may be put in the form

$$1+\frac{3^{32\cdot 37}}{3}+\frac{5^{32\cdot 37}}{5}+\cdots+\frac{35^{32\cdot 37}}{35}$$

and $2^{58\cdot 67}$ was reduced modulo 67^2 so that $2^{58\cdot 67}\equiv 3188 \pmod{67^2}$. Similarly $3^{58\cdot 67}\equiv 3859$, and it was noted that $5^{58\cdot 67}\equiv (3^2\cdot 2^3)^{58\cdot 67} \pmod{67^2}$, so that $5^{58\cdot 67}$ is obtained easily from our reductions for the cases 2 and 3. Similarly $7^{58\cdot 67}\equiv (5\cdot 2^3\cdot 3)^{58\cdot 67} \pmod{67^2}$ and so on. These powers of all the odd integers less than $l-1$ were found modulo 67^2, then the expressions $1/3$, $1/5$, $\cdots$, $1/65$ were computed modulo 67^2, and the resulting integers were multiplied in the corresponding powers in the formula.

Applying the above mentioned computations to Theorem II, we establish Fermat's last theorem for all the regular primes less than 211 excepting 157.

Theorem III includes two assumptions which are satisfied simultaneously only in the cases $l=37$ and $l=101$.

In order to test Theorem IV for the known irregular primes less than 211 the symbol

$$\left\{\frac{E_a}{\mathfrak{p}}\right\}$$

was computed for the particular values of l and a. For $l=37$ we have $a=16$, $p=149$, $r=2$, and $\zeta\equiv 17 \pmod{\mathfrak{p}}$. Hence to test whether or not the above

mentioned symbol is 1 we reduce $E_{16}(17)$ modulo 149, obtain its index modulo 148 using Cahen's tables of indices in the second volume of his *Théorie des Nombres* (which companion tables were checked independently by comparing one with the other). These computations were carried out by Professor Stafford (except for $l=157$, which case was computed by the writer) and are fully described in the joint article already referred to. She found ind $E_{16}(17)\equiv 24 \pmod{37}$ which agrees with Kummer's results in his 1857 Memoir. Similarly for $l=59$, $p=709$, $\zeta\equiv 385 \pmod{\mathfrak{p}}$, $r=2$, $a=22$, and ind $E_{22}(385)\equiv 50 \pmod{59}$. For $l=67$, $p=269$, $\zeta\equiv 47 \pmod{\mathfrak{p}}$, $r=2$, $a=29$, and ind $E_{29}(47)\equiv 4 \pmod{67}$. The cases 59 and 67 agree with Kummer's computations in his Memoir just cited. For $l=101$, $p=809$, $\zeta\equiv 100 \pmod{\mathfrak{p}}$, $r=2$, $a=34$, and ind $E_{34}(100)\equiv 45 \pmod{101}$. For $l=103$, $p=619$, $\zeta\equiv 315 \pmod{\mathfrak{p}}$, $r=5$, $a=12$, and ind $E_{12}(315)\equiv 65 \pmod{103}$. For $l=131$, $p=263$, $r=2$, $\zeta\equiv 100 \pmod{131}$, $a=11$ and ind $E_{11}(100)\equiv 52 \pmod{131}$. For $l=149$, $p=1093$, $r=10$, $\zeta\equiv 354 \pmod{\mathfrak{p}}$ and $a=65$, we have $E_{65}(354)\equiv 127 \pmod{149}$. For $l=157$, $p=1571$, $r=139$, $\zeta\equiv 1024 \pmod{\mathfrak{p}}$, and $a=31$, we have ind $E_{31}(1024)\equiv 150 \pmod{157}$. For $l=157$, $a=55$, we have $p=1571$, $r=139$, $\zeta\equiv 1024 \pmod{\mathfrak{p}}$ and ind $E_{55}\equiv 39 \pmod{157}$. These computations show that the assumptions in Theorem IV are satisfied for all the known irregular primes we are considering, and also show that the second factor of the class number of each of the fields defined by these known irregular primes is prime to l. The latter statement follows because if the second factor of the cyclotomic field defined by l is divisible by l, then $E_{a_1}^{n_1} E_{a_2}^{n_2} \cdots E_{a_e}^{n_e}$ is the lth power of a unit in the field with not all the n's divisible by l. It follows therefrom that each E_a^n is the lth power of a unit in the field.* Through these results and those obtained concerning the A's and B's, we have also a test of Theorem I for all the irregular primes.

We have then tested the first four theorems for all of our known irregular primes and we may enunciate

THEOREM VII. *The equation*

$$x^n + y^n = z^n$$

is impossible in rational integers none zero if

$$2 < n < 211.$$

Note. Since the above was written, Theorem VII has been extended to $n<269$.

* Vandiver, Bulletin of the American Mathematical Society, vol. 35 (1929) pp. 323–335.

UNIVERSITY OF TEXAS,
AUSTIN, TEXAS

THE DEGREE AND CLASS OF MULTIPLY TRANSITIVE GROUPS, II*

BY

W. A. MANNING

1. The paper to which this is a sequel had to do only with those multiply transitive groups of class $\mu(>3)$ in which at least one substitution of degree μ is of even order.† Among other results, it was there proved that the degree n of triply transitive groups of class $\mu(>3)$, which contain a substitution of even order on μ letters, does not exceed 2μ. Triply transitive groups of class μ and degree 2μ exist. This measure of success was due to the simplicity of structure of diedral rotation groups of class μ generated by two similar substitutions of degree μ and order 2. Up to the present time no better limit than that of Bochert, $n \leqq 3\mu(\mu>6)$, has been found to apply to these triply transitive groups of class $\mu(>3)$ in which all substitutions of degree μ are of odd order.‡ But it will here be proved that if such a group contains a substitution of order p^e (p an odd prime) on μ letters,

$$\mu > \frac{n}{2}\left(1 - \frac{1}{p^e}\right) - \frac{2}{p^e}.$$

For doubly transitive groups the lower limit for the class $\mu(>3)$ in terms of the degree when there is a substitution of even order on μ letters present in the group is given by

$$\mu > \frac{n}{2} - \frac{n^{1/2}}{2} - 1.\S$$

Whenever applicable, this replaces Bochert's limit of $n/3 - 2n^{1/2}/3$.¶ We can now prove that, when a doubly transitive group of class $\mu(>3)$ contains a substitution of odd prime order p, and (when $p=3$) n is sufficiently large,

$$\mu > \frac{n}{2}\left(1 - \frac{1}{p}\right) - \frac{n^{1/2}}{2}\left(1 - \frac{1}{p^2}\right)^{1/2} - 1.$$

---

* Presented to the Society, triply and quadruply transitive groups, San Francisco Section, October 27, 1928; doubly transitive groups, December 31, 1928. Received by the editors in December, 1928.

† Manning, these Transactions, vol. 18 (1917), p. 463.

‡ Bochert, Mathematische Annalen, vol. 40 (1892), p. 185.

§ Manning, Bulletin of the American Mathematical Society, vol. 20 (1914), p. 468.

¶ Bochert, Mathematische Annalen, vol. 49 (1897), p. 144.

For $p=3$, this is only $\mu>n/3-(2n)^{1/2}/3-1$. It was stated by Bochert without proof that $n/3-(2n)^{1/2}/3$ is a limit to which his ascending series of inferior limits for μ approaches on repeated use of his inequality (19). That is a mistake. His inequality is satisfied by $n/3-(2n)^{1/2}/3$ and also by $n/3-(2n)^{1/2}/3-1/2$ while a true inferior limit for μ is a number that does not satisfy the given inequality and which therefore is less than $n/3-(2n)^{1/2}-1/2$.

2. Bochert's Lemma will be used in the following form:

If the substitutions S and T have exactly m letters in common, and if S replaces q, and T r, common letters by common letters, the degree of $S^{-1}T^{-1}ST$ is not greater than $3m-q-r$.

This lemma remains indispensable in studying quadruply transitive groups of class $\mu(>3)$ and those doubly and triply transitive groups of class $\mu(>3)$ in which all the substitutions of degree μ are of order 3. But additional information is given in other cases by the following lemma:

LEMMA. *If S and T are two substitutions of degree μ and odd order d which generate a group of class μ, and if no power of S is commutative with a power of T (identity excepted), S and T have at least $\mu/2-\mu/(2d)$ letters in common.*

Let S and T have exactly m letters in common, m roman letters, say. Let the other letters of S and T be greek letters. From Bochert's Lemma it is known that $m\geqq[\mu/3]$, the integral part of $\mu/3$. Then no proof is needed when $d=3$. We assume that $d\geqq 5$. Now we say that S has s_i cycles each of which contains i roman letters, and that T has t_j cycles each of which contains j romans. Then

$$s_0+s_1+\cdots+s_d=t_0+t_1+\cdots+t_d=\frac{\mu}{d},$$

the number of cycles in S and T. Also

$$\begin{aligned}s_1+2s_2+3s_3+\cdots+ds_d&=t_1+2t_2+3t_3+\cdots+dt_d\\&=\mu/3-\epsilon+k_1=\mu/3+k,\end{aligned}$$

where $\epsilon=0$, $1/3$, or $2/3$, as required to make $\mu/3-\epsilon=[\mu/3]$, and k_1 is a positive integer or zero.

Consider a cycle of S in which there are i romans. It generates a regular cyclic group. Therefore in the $d-1$ powers of this cycle every sequence of two letters occurs once and only once; then in these $d-1$ powers of one cycle of S there are $i(i-1)$ roman sequences. In the $d-1$ powers of S the number of roman sequences is exactly $2s_2+6s_3+\cdots+i(i-1)s_i+\cdots+d(d-1)s_d$, and the average number of roman sequences in the $d-1$ powers of S is $[2s_2+6s_3+\cdots+d(d-1)s_d]/(d-1)$. Similarly, in the powers of T the

average number of roman sequences is $[2t_2+6t_3+\cdots+d(d-1)t_d]/(d-1)$. From the powers of S and T let two substitutions S^u and T^v be chosen, each of which has the average or more than the average number of roman sequences. Since the number of roman letters is $\mu/3+k$, the total number of roman sequences in S^u and T^v jointly cannot exceed $3k$. For Bochert's Lemma asserts that the degree of $S^{-u}T^{-v}S^uT^v$ is at most $\mu+3k$ diminished by the number of roman sequences in S^u and T^v. And since S^u and T^v are not commutative this degree is at least μ. Then

$$\sum_{i=0}^{d}\frac{i(i-1)s_i}{d-1}+\sum_{j=0}^{d}\frac{j(j-1)t_j}{d-1}\leqq 3k,$$

and one of the two summation terms of this inequality is not greater than $3k/2$, say the first. Thus we have simultaneously

$$s_0+s_1+s_2+\cdots+s_d=\frac{\mu}{d},$$

$$s_1+2s_2+3s_3+\cdots+ds_d=\frac{\mu}{3}+k,$$

$$2s_2+6s_3+12s_4+\cdots+d(d-1)s_d\leqq\frac{3}{2}(d-1)k.$$

Let two terms, s_x and s_y, be eliminated $(x<y)$. The result is

$$\frac{(x+y-1)d-3xy}{3d}\mu+\sum_{i=0}^{d}(i-x)(i-y)s_i\leqq\frac{1}{2}(3d-2x-2y-1)k.$$

If now $x+1$ be put for y,

$$\sum_{i=0}^{d}(i-x)(i-x-1)s_i\geqq 0,$$

and therefore

$$\mu\leqq\frac{3d(3d-4x-3)}{2x(2d-3x-3)}k.$$

We may choose the integer x to make the coefficient of k a minimum. The continuous curve

$$y=\frac{3d(3d-4x-3)}{2x(2d-3x-3)}$$

has a minimum between $x=0$ and $x=2d/3-1$. From $D_xy=0$ we get $x=$

$\{3d-3-[(d+1)\ (d+3)]^{1/2}\}/4$, lying between $(d-3)/2$ and $(d-1)/2$, and to these two integers correspond equal ordinates, $6d/(d-3)$. Then

$$k \geqq \frac{d-3}{6d}\mu,$$

and

$$m \geqq \frac{d-1}{2d}\mu.$$

It follows that the degree of $\{S, T\}$ is at most $3\mu/2+\mu/(2d)$.

3. We prove the following theorem:

THEOREM I. *The class $\mu(>3)$ of a triply transitive group of degree n that contains a substitution of degree μ and of order p^c (p an odd prime) is greater than*

$$\frac{n}{2}\left(1-\frac{1}{p^c}\right)-\frac{2}{p^c}.$$

One of the substitutions of degree μ in the group G is $S=(a\cdots b\cdots)\cdots$, of order p^c. In case the exponent c is greater than unity, b is one of the letters in the same cycle of $S^{p^{c-1}}$ as a, and therefore in the same cycle as a in every power of S. There are in G subsitutions similar to S which displace a and fix b. Let $S_1, S_2, \cdots, S_w$ be a complete set of w such substitutions, conjugate under the transitive subgroup $G(a)(b)$ of G that fixes a and b. No power of S_i $(i=1, 2, \cdots, w)$ is commutative with a power of S. The substitutions of this complete set of conjugates displace exactly

$$w+\frac{w(\mu-1)(\mu-2)}{n-2}$$

letters of S.* For they all displace a, and each of the other $w(\mu-1)$ letters in the set $S_1, S_2, \cdots, S_w$ occurs as often as any other, that is, $w(\mu-1)/(n-2)$ times. Thus the $\mu-2$ letters of S, disregarding a and b, occur in the set $S_1, S_2, \cdots$ the number of times stated above. Then, by our lemma,

$$w+\frac{w(\mu-1)(\mu-2)}{n-2} \geqq w\left(\frac{\mu}{2}-\frac{\mu}{2p^c}\right),$$

whence

$$\frac{n-\mu}{p^c}-(\mu-2)\left(\frac{n}{2}-\frac{n}{2p^c}-\mu\right) \geqq 0.$$

* For a detailed explanation of this and of similar formulas which follow, see Bochert, Mathematische Annalen, vol. 40 (1892), p. 176.

If now we put $n/2-n/(2p^c)-\delta$ for μ in the left member of this inequality, we get

$$-\frac{n}{2}\left(\delta-\frac{\delta}{p^c}-\frac{1}{p^c}-\frac{1}{2p^{2c}}\right)+\delta^2+2\delta-\frac{\delta}{p^c}.$$

This is positive if $\delta=1/p^c$, but if $\delta=2/p^c$ it reduces to

$$-\frac{n}{2}\left(\frac{1}{p^c}-\frac{3}{p^{2c}}\right)+\frac{4}{p^c}+\frac{2}{p^{2c}},$$

negative if n is greater than $8+28/(p^c-3)$. It is known that there is no triply transitive group of class 5. Hence the theorem is true as stated without exception.

4. We prove the following theorem:

THEOREM II. *The class $\mu(>3)$ of a triply transitive group of degree n in which every substitution of degree μ is of order 3 is not less than $(n+4)/3$.*

Let $S=(abc)\cdots$, a substitution of degree μ of G. Let $S_1=(a\cdots)\cdots(b)$, similar to S. Under $G(a)(b)$, S_1 is one of w conjugates, no one of which is commutative with S. If S and S_i have exactly $\mu/3$ common letters (their roman letters), S_i can have no roman sequence. This fact is an immediate consequence of Bochert's Lemma. In other words, if S_i has a roman sequence, it has at least $\mu/3+1$ letters in common with S. If S and S_i have $\mu/3+1$ letters in common, S, because it has only $\mu/3$ cycles, has some two letters of S_i in sequence. Hence, if S_i has a cycle of three roman letters, which means three roman sequences, S and S_i have $\mu/3+2$ or more common letters. Now in $w/(n-2)$ substitutions of the set $S_1, S_2, \cdots, S_w$ the letter a is followed by a given letter of $G(a)(b)$. Then in $w(\mu-2)/(n-2)$ substitutions S_i, a is followed by a letter of S, and similarly a is preceded by a letter of S in $w(\mu-2)/(n-2)$ substitutions S_i. Therefore

$$w+\frac{w(\mu-1)(\mu-2)}{n-2}\geqq\frac{w\mu}{3}+2w\frac{\mu-2}{n-2}.$$

This inequality reduces to

$$\frac{(\mu-2)(\mu-3)}{n-2}\geqq\frac{\mu-3}{3},$$

whence, because by hypothesis $\mu>3$,

$$n\leqq 3\mu-4.$$

5. We prove the following theorem:

THEOREM III. *The degree of a quadruply transitive group of class* μ (>3) *does not exceed* $2\mu+1$.

For quadruply transitive groups of class μ Bochert's limit is $2\mu+2$, while the limit found in the preceding paper under the restriction that one of the substitutions of degree μ is of even order is $2\mu-1$. In the following proof it is assumed that all the substitutions of degree μ in G are of odd order.

Two cases are to be distinguished: (1) At least one substitution of degree μ is of order >3. (2) All substitutions of degree μ are of order 3.

Case 1. Let $S=(a\cdots b\cdots)\cdots$ be a substitution of degree μ and of order >3. The letters a and b are in the same cycle of S but are not adjacent. There are w substitutions $S_1=(b)\ (a\cdots)\cdots$, $S_2, \cdots$, conjugate under $G(a)(b)$, a doubly transitive subgroup of G. It was shown in §3 that the number of times the letters of S are to be found in the set $S_1, S_2, \cdots$ is $w+w(\mu-1)(\mu-2)/(n-2)$. The number of times letters of S occur just before or just after a in the set $S_1, S_2, \cdots$ is $2w(\mu-2)/(n-2)$. The number of sequences in $S_1, S_2, \cdots$ is $w(\mu-2)$, if we exclude those of which one letter is a. Since $G(a)(b)$ is doubly transitive each sequence occurs as often as any other and therefore exactly $w(\mu-2)/[(n-2)(n-3)]$ times. Now there are $(\mu-2)(\mu-3)$ permutations two at a time of the letters of S (excluding a and b). Therefore $2w(\mu-2)/(n-2)+w(\mu-2)^2(\mu-3)/[(n-2)(n-3)]$ is the number of sequences in $S_1, S_2, \cdots$ of which both letters are letters of S. The number of times the letter before a (and the letter after a) in S occurs in the set $S_1, S_2, \cdots$ is $2w(\mu-1)/(n-2)$. Excluding a and b, S has $\mu-4$ sequences. Excluding a, S_i has $(\mu-1)(\mu-2)$ permutations of letters two at a time, and therefore the set $S_1, S_2, \cdots$ contains $w(\mu-1)(\mu-2)$ such permutations of $\mu-1$ letters two at a time, each occurring $w(\mu-1)(\mu-2)\cdot[(n-2)(n-3)]^{-1}$ times, so that each sequence of S (not including a or b) is in $w(\mu-1)(\mu-2)/[(n-2)(n-3)]$ subsitutions of the set $S_1, S_2, \cdots$. Therefore $2w(\mu-1)/(n-2)+w(\mu-1)(\mu-2)(\mu-4)/[(n-2)(n-3)]$ is the number of times permutations two at a time of letters of $S_1, S_2, \cdots$ are a sequence in S. Therefore the sum of the degrees of the w commutators $S^{-1}S_i^{-1}SS_i$ $(i=1, 2, \cdots, w)$ gives

$$3w+\frac{3w(\mu-1)(\mu-2)}{n-2}-\frac{2w(\mu-2)}{n-2}-\frac{w(\mu-2)^2(\mu-3)}{(n-2)(n-3)}$$
$$-\frac{2w(\mu-1)}{n-2}-\frac{w(\mu-1)(\mu-2)(\mu-4)}{(n-2)(n-3)}\geqq w\mu.$$

Whence, if z is put for $n-2\mu-3$,

$$(\mu - 3)[z^2 + (\mu + 5)z] + 2(\mu^2 - 10) \leqq 0,$$

or

$$z^2 + (\mu + 5)z + 2\mu + 6 \leqq 0 \quad (\mu > 5).$$

Therefore

$$n - 2\mu - 3 \leqq -2, \quad n \leqq 2\mu + 1.$$

Case 2. The substitutions of degree μ are all of order 3. We have $S = (abc) \cdots$ and the set of w conjugates under $G(a)(b)$: $S_1 = (b)(a \cdots) \cdots$, $S_2, \cdots$. Since $G(a)(b)$ is doubly transitive we can choose the two letters in one of these substitutions, in S_1 say, at will from the $n-2$ letters of $G(a)(b)$. If $S_1 = (ac\alpha)(b) \cdots$, $S^{-1}S_1^{-1}SS_1 = (c\alpha)(ab) \cdots$. If this substitution is of degree μ, $n \leqq 2\mu - 1$, and the theorem is true. Hence the degree of this commutator is not less than $\mu+1$. Also if $S_1 = (a\alpha c)(b) \cdots$, $S^{-1}S_1^{-1}SS_1 = (ac)(b\alpha) \cdots$, which must be of degree $\geqq \mu+1$.

In the first cycle of each of the substitutions $S_1, S_2, \cdots$ there occurs every possible sequence of the letters of $G(a)(b)$, and any such sequence occurs $w/[(n-2)(n-3)]$ times. Then the $2(n-\mu)$ sequences in which the letter c of S is followed or preceded by one of the $n-\mu$ letters of G fixed by S occur $2w(n-\mu)/[(n-2)(n-3)]$ times in the first cycle of the substitutions $S_1, S_2, \cdots$. Thus the commutator of S and $2w(n-\mu)/[(n-2)(n-3)]$ substitutions of the set $S_1, S_2, \cdots$ is of degree $\geqq \mu+1$.

In the set $S_1, S_2, \cdots$, a is a letter of $2w(\mu-2)/(n-2)$ sequences in which the other letter is in S, and c occurs in $w(\mu-1)/(n-2)$ of the substitutions $S_1, S_2, \cdots$. In the set $S_1, S_2, \cdots$ there are $w(\mu-2)^2(\mu-3)/[(n-2)(n-3)]$ sequences, not involving a, whose two letters are the $\mu-2$ letters $c, d, \cdots$ of S. There are $\mu-3$ sequences in S (not including a or b) and each such permutation of letters two at a time occurs in $S_1, S_2, \cdots$, $w(\mu-1)(\mu-2) \cdot [(n-2)(n-3)]^{-1}$ times. Hence

$$3w + \frac{3w(\mu-1)(\mu-2)}{n-2} - \frac{2w(\mu-2)}{n-2} - \frac{w(\mu-1)}{n-2} - \frac{w(\mu-2)^2(\mu-3)}{(n-2)(n-3)} - \frac{w(\mu-1)(\mu-2)(\mu-3)}{(n-2)(n-3)} \geqq w\mu + \frac{2w(n-\mu)}{(n-2)(n-3)},$$

or, with $z = n - 2\mu - 3$,

$$z^2 + (\mu + 4)z + \mu + 4 \leqq 0,$$

and therefore

$$n \leqq 2\mu + 1.$$

6. We prove the following theorem:

THEOREM IV. *Let μ be the class and n the degree of a doubly transitive group in which one of the substitutions of degree $\mu(>3)$ is of prime order $p(>3)$; then*

$$\mu > \frac{n}{2}\left(1-\frac{1}{p}\right) - \frac{n^{1/2}}{2}\left(1-\frac{1}{p^2}\right)^{1/2} - 1.$$

There is a substitution $S=(ab\cdots)\cdots$ of prime order p (>3) in the given doubly transitive group G, and S is one of w conjugates under G. Any sequence of two letters, as ab, occurs $w\mu/[n(n-1)]$ times in this set. The number of possible sequences in n letters, in which one letter belongs to S and the other does not, is $2\mu(n-\mu)$. Then in the complete set of w conjugate substitutions such sequences as $a\alpha$ and αa occur in all $2w\mu^2(n-\mu)\cdot[n(n-1)]^{-1}$ times. There can be at most $p-1$ such sequences in a cycle. Let us say that $S_1, S_2, \cdots, S_y$ are y substitutions of the set not commutative with S. Then

$$y \geqq \frac{2w\mu(n-\mu)}{n(n-1)(1-1/p)}.$$

We recall that if one of the w conjugate substitutions has exactly x letters in common with S,

$$\sum_w x = \frac{w\mu^2}{n},$$

$$\sum_w x(x-1) = \frac{w\mu^2(\mu-1)^2}{n(n-1)},$$

and

$$\sum_w \left(x-\frac{\mu^2}{n}\right)^2 = \frac{w\mu^2(n-\mu)^2}{n^2(n-1)}.$$

Now by our lemma, if S and S_i $(i=1, 2, \cdots, y)$ have x_i letters in common,

$$x_i \geqq \frac{\mu}{2}\left(1-\frac{1}{p}\right),$$

and therefore

$$\sum_{i=1}^{y} x_i \geqq \frac{y\mu}{2}\left(1-\frac{1}{p}\right).$$

Let us now restrict our attention to those doubly transitive groups for which

$$\mu \leqq \frac{n}{2}\left(1-\frac{1}{p}\right).$$

For such groups

$$\frac{\mu}{2}\left(1-\frac{1}{p}\right) - \frac{\mu}{2}\,\frac{2\mu}{n} \geqq 0,$$

and therefore

$$\sum_{i=1}^{y}\left(x_i-\frac{\mu^2}{n}\right)^2 \geqq y\left(\frac{\mu}{2}-\frac{\mu}{2p}-\frac{\mu^2}{n}\right)^2.$$

Next,

$$\sum_{w-y}\left(x-\frac{\mu^2}{n}\right)^2 \leqq \frac{w\mu^2(n-\mu)^2}{n^2(n-1)}-y\left(\frac{\mu}{2}-\frac{\mu}{2p}-\frac{\mu^2}{n}\right)^2.$$

Also

$$\sum_{w-y}\left(x-\frac{\mu^2}{n}\right)^2 \geqq \frac{1}{w-y}\left[\sum_{w-y}\left(x-\frac{\mu^2}{n}\right)\right]^2$$

$$\geqq \frac{1}{w-y}\left(\frac{y\mu}{2}-\frac{y\mu}{2p}-\frac{y\mu^2}{n}\right)^2,$$

so that

$$\frac{w\mu^2(n-\mu)^2}{n^2(n-1)}-y\left(\frac{\mu}{2}-\frac{\mu}{2p}-\frac{\mu^2}{n}\right)^2 \geqq \frac{y^2}{w-y}\left(\frac{\mu}{2}-\frac{\mu}{2p}-\frac{\mu^2}{n}\right)^2,$$

or

$$\frac{w\mu^2(n-\mu)^2}{n^2(n-1)}-\frac{wy}{w-y}\left(\frac{\mu}{2}-\frac{\mu}{2p}-\frac{\mu^2}{n}\right)^2 \geqq 0,$$

or

$$\left(\frac{w}{y}-1\right)\frac{\mu^2(n-\mu)^2}{n^2(n-1)}-\left(\frac{\mu}{2}-\frac{\mu}{2p}-\frac{\mu^2}{n}\right)^2 \geqq 0.$$

Substituting for w/y,

$$\frac{\mu(n-\mu)(1-1/p)}{2n}-\frac{\mu^2(n-\mu)^2}{n^2(n-1)}-\frac{\mu^2}{n^2}\left(\frac{n}{2}-\frac{n}{2p}-\mu\right)^2 \geqq 0,$$

or finally,

$$\frac{n(n-\mu)(1-1/p)}{2\mu}-\frac{(n-\mu)^2}{n-1}-\left(\frac{n}{2}-\frac{n}{2p}-\mu\right)^2 \geqq 0.$$

Let $n/2-n/(2p)-\epsilon n^{1/2}/2-\delta$ be substituted for μ in the final inequality. In the resulting polynomial in $n^{1/2}$ it will be seen that the coefficient of n^3 vanishes if $\epsilon=(1-1/p^2)^{1/2}$. With this value of ϵ agreed upon, the polynomial may be written

$$\left(2+\frac{2}{p^2}-4\delta+\frac{4\delta}{p}\right)\epsilon n^{5/2}+\left(\frac{2\epsilon^2}{p}+8\delta-4\delta^2+\frac{4\delta^2}{p}\right)n^2$$

$$+\left(8\delta+\frac{8\delta}{p}+12\delta^2-2+\frac{2}{p}\right)\epsilon n^{3/2}+\left(8\delta^2+8\delta+\frac{8\delta}{p}-4+\frac{4}{p}\right)\delta n.$$

If $\delta=\frac{1}{2}$, this polynomial is positive; if $\delta=1$, it reduces to

$$-\left(1-\frac{2}{p}-\frac{1}{p^2}\right)\epsilon n^{3/2}+\left(2+\frac{3}{p}-\frac{1}{p^3}\right)n+\left(9+\frac{5}{p}\right)\epsilon n^{1/2}+6+\frac{6}{p},$$

which is negative for $n>57$. Any possible cases of exception are covered by the known theorems on non-alternating primitive groups which contain substitutions of order p and degree qp $(q=1, 2, 3, 4)$.*

7. The above proof is valid for $p=3$ until we come to the discussion of the polynomial

$$-n^{3/2}+10\cdot 2^{1/2}n+48n^{1/2}+27\cdot 2^{1/2},$$

which is negative when $n>292$. If however we put $\delta=4/3$, the polynomial is

$$-3n^{3/2}+11\cdot 2^{1/2}n+77n^{1/2}+58\cdot 2^{1/2},$$

and this is negative if $n>73$. If $n=73, 70, 69, 60, 59$, $[n-(2n)^{1/2}-4]/3=$ 18.9, 18.1, 17.8, 15.0, 14.7, respectively. It can be shown that this limit holds for $n\leqq 73$. It is known to be true for doubly transitive groups of class 6, 9, and 12†. For groups of class 15 or 18, we can use the following theorem:

A primitive group that contains a substitution of order p and degree pq (p an odd prime, $p<q<2p+3$), contains a transitive subgroup the degree of which is not greater than the larger of the two numbers $pq+q^2-q$ and $2q^2-p^2$.‡

Thus our doubly transitive group of class 18 contains a transitive subgroup H of degree $\leqq 63$, and we know from the proof of the theorem cited that the latter is generated by substitutions of order 3 and degree 18. We are concerned only with $n=70, 71, 72, 73$. Then G is more than doubly transitive unless the transitive subgroup H is imprimitive. Since H is generated by similar substitutions of order 3 and degree 18, its systems of imprimitivity are of two, three, or six letters. If H has systems of imprimitivity of six letters each, its degree is not greater than 60, and G has a doubly transitive subgroup H' of degree $\leqq 67$; and if the systems are of two or three letters only, the same is true.§ Thus G is more than doubly transitive and we should have (by §4) $\mu>n/3$.

If G is of class 15, H is of degree $\leqq 41$, and the doubly transitive subgroup H' is of degree $\leqq 46$.

---

* Manning, these Transactions, vol. 10 (1909), p. 247.

† Manning, these Transactions, vol. 6 (1905), p. 45; American Journal of Mathematics, vol. 35 (1913), p. 229.

‡ Manning, these Transactions, vol. 12 (1911), p. 382, §12.

§ Manning, these Transactions, vol. 7 (1906), p. 499; *Primitive Groups*, part 1, 1921, p. 93.

Our result for $p=3$ leaves much to be desired but is at any rate of the same form as Theorem IV:

THEOREM V. *Let μ be the class and n (>292) the degree of a doubly transitive group in which one of the substitutions of degree $\mu(>3)$ is of order* 3; *then*

$$\mu > \frac{n}{3} - \frac{(2n)^{1/2}}{3} - 1.$$

If $n \leqq 292$,

$$\mu > \frac{n}{3} - \frac{(2n)^{1/2}}{3} - \frac{4}{3}.$$

STANFORD UNIVERSITY,
PALO ALTO, CALIF.

ALGEBRAIC COMBINATIONS OF EXPONENTIALS*

BY

J. F. RITT

By an *exponential polynomial*, we shall mean a function

$$P(z) = a_0 e^{\alpha_0 z} + \cdots + a_m e^{\alpha_m z}$$

with arbitrary complex numbers for a's and α's. We shall refer to the α's as the *exponents* of $P(z)$.

We study here functions defined by an equation

$$P_n w^n + P_{n-1} w^{n-1} + \cdots + P_0 = 0, \tag{1}$$

with every P an exponential polynomial. When the exponents of the P's are all integers, the substitution $e^z = u$ converts w into an algebraic function of u. Thus, in a sense, the theory of the equation (1) is a generalization of the theory of algebraic functions.

The present investigation arose out of the problem of determining all *uniform* functions which satisfy an equation (1). It is a natural conjecture that if w, defined by (1), is uniform, then $w = Q/R$ with Q and R exponential polynomials. One might, further, expect the exponents in Q and R to be linear combinations of the exponents in the P's, with rational coefficients. Both of these conjectures are verified by our work.

We have already proved† that if the quotient of two exponential polynomials is an integral function, the quotient is an exponential polynomial. This result is obtained again here, under the weaker hypothesis that the quotient is analytic in a sector of opening greater than π.

The problem of uniformity settled, we are able to discuss the relation between the reducibility properties of (1) and the number of analytic functions which (1) defines. Our result, which generalizes a well known theorem on algebraic functions, is stated in §21.

In examining these questions, we were led to study the behavior, for large values of z, of functions determined by (1). It turns out that the Riemann surface for w can be divided into a finite number of sectors, in each of which, after a border of the sector of finite width is removed, w can be represented by a Dirichlet series with *complex* exponents. The sectors admit of precise description.

* Presented to the Society, February 23, 1929; received by the editors May 1, 1929.

† Ritt, (1). (List of references at end of paper.)

In §21 we prove that if S is any sector of opening greater than π, the branch points of w lying in S, and lying outside an arbitrarily large circle, serve, in their totality, to connect up all of the branches of w, and, in fact, to generate the entire group of monodromy of w.

The developability of w in Dirichlet series with complex exponents is proved by the Newton polygon method of algebraic function theory, with appropriate modifications. We have found serviceable, also, the results of Tamarkin, Wilder and Pólya on the zeros of exponential polynomials, Hille's work on Dirichlet series with complex exponents, and the notion of almost periodic function.

I. Equations whose coefficients have real exponents

1. We consider an equation*

$$f(w) = P_n(z)w^n + P_{n-1}(z)w^{n-1} + \cdots + P_1(z)w + P_0(z) = 0, \tag{2}$$

where each P_i is a function defined by a Dirichlet series with a half-plane of *absolute* convergence

$$P_i(z) = a_{i1}e^{\lambda_1 z} + a_{i2}e^{\lambda_2 z} + \cdots + a_{ip}e^{\lambda_p z} + \cdots,$$

the λ's, independent of i, being real, and such that

$$\lambda_1 > \lambda_2 > \cdots > \lambda_p > \cdots ; \quad \lim_{p\to\infty} \lambda_p = -\infty.$$

We make the non-restrictive assumption that $P_0 \not\equiv 0, P_n \not\equiv 0$.

2. We shall prove that a real A exists, such that, for $x > A$ $(z = x + y(-1)^{1/2})$, the n solutions $w_1, \cdots, w_n$ of (2) are analytic and representable by absolutely convergent Dirichlet series

$$w_i(z) = c_{i1}e^{\rho_1 z} + c_{i2}e^{\rho_2 z} + \cdots + c_{ip}e^{\rho_p z} + \cdots, \tag{3}$$

the ρ's, which decrease toward $-\infty$, having expressions

$$\rho_p = \frac{1}{r}(m_{p1}\lambda_1 + m_{p2}\lambda_2 + \cdots + m_{pq_p}\lambda_{q_p}), \tag{4}$$

with r a positive integer independent of p, and with integral m's.

3. It will suffice to show that (2) has *one* solution of the form (3).

If P_i $(i = 0, \cdots, n)$ is not identically zero, we represent by σ_i the largest λ which occurs in P_i in a term with a coefficient a not zero. If $P_i \equiv 0$, we let $\sigma_i = -\infty$.

---

* It is convenient to write $f(w)$ rather than $f(w, z)$.

Let g_1 be the greatest value of i $(i=1, \cdots, n)$ which makes

$$\frac{\sigma_0 - \sigma_i}{i} \tag{5}$$

a minimum. Let ρ_1 denote this minimum.

Let $\alpha_i^{(1)}$ $(i=1, \cdots, g_1)$ denote the coefficient of $e^{\sigma_i z}$ in P_i, or denote 0, according as (5) equals ρ_1 or exceeds ρ_1. Let $\alpha_0^{(1)}$ be the coefficient of $e^{\sigma_0 z}$ in P_0.

Let c_1 be any root of the equation

$$\phi_1(c) = \alpha_0^{(1)} + \alpha_1^{(1)} c + \cdots + \alpha_{g_1}^{(1)} c^{g_1} = 0. \tag{6}$$

We put $w = c_1 e^{\rho_1 z} + w'$. Then (2) gives an equation for w',

$$f_1(w') = P_n'(z) w'^n + \cdots + P_1'(z) w' + P_0'(z) = 0, \tag{7}$$

where $P_n' = P_n$, and where every P' is a function represented by a Dirichlet series with a half-plane of absolute convergence, whose exponents are the λ's plus integral multiples of ρ_1. The exponents in the P''s will have expressions like the second member of (4).

If $P_0' \equiv 0$, the function $w = c_1 e^{\rho_1 z}$ is a solution of (2). If $P_0' \not\equiv 0$, we treat $f_1(w')$ as we did $f(w)$, determining a ρ_2 and a c_2, then making the substitution $w' = c_2 e^{\rho_2 z} + w''$.

We build, in this way, a series

$$c_1 e^{\rho_1 z} + c_2 e^{\rho_2 z} + c_3 e^{\rho_3 z} + \cdots . \tag{8}$$

If, at some stage, a $P_0^{(j)}$ is found which is identically zero, the series (8) is finite and is a solution of (2), of type (3).

4. We assume, in what follows, that (8) is an infinite series. We say that

(a) $$\rho_1 > \rho_2 > \rho_3 > \cdots ; \quad \lim_{p \to \infty} \rho_p = -\infty .$$

(b) *The ρ's are of the form* (4).

(c) *The series* (8) *has a half-plane of absolute convergence.*

(d) *The function defined by* (8) *satisfies* (2).

We remark that (b) is obvious on the basis of §3.

5. The equation which defines c_2 in (8) is

$$\phi_2(c) = \alpha_0^{(2)} + \cdots + \alpha_{g_2}^{(2)} c^{g_2} = 0. \tag{9}$$

We are going to show that* if c_1 is a root of (6) of multiplicity s_1, then

* In §5, we follow closely Hensel und Landsberg, *Algebraische Funktionen*, Leipzig, 1902, pp. 53 et seq. In §§6, 7, however, situations peculiar to the Dirichlet series problem require us to deviate from the methods used in algebraic function theory.

$g_2 \leqq s_1$. We prove also that $\rho_{p+1} < \rho_p$ for every p.

For c arbitrary, we have

$$f(ce^{\rho_1 z}) = \phi_1(c)e^{\sigma_0 z} + \psi_1(c,z)e^{\tau z}, \tag{10}$$

where $\psi_1(c, z)$ is a polynomial in c with coefficients which are Dirichlet series with non-positive exponents, and where $\tau < \sigma_0$.

If, in (10), we put $c = c_1 + w'e^{-\rho_1 z}$, the first member becomes $f(c_1 e^{\rho_1 z} + w')$ which is $f_1(w')$. Hence

$$f_1(w') = \phi_1(c_1 + w'e^{-\rho_1 z})e^{\sigma_0 z} + \psi_1(c_1 + w'e^{-\rho_1 z}, z)e^{\tau z}. \tag{11}$$

Comparing this result with (7), we find that

$$P_i' = e^{(\sigma_0 - i\rho_1)z}\frac{\phi_1^{(i)}(c_1)}{i!} + e^{(\tau - i\rho_1)z}\frac{\psi_1^{(i)}(c_1, z)}{i!} \tag{12}$$

where the superscript (i) indicates i differentiations with respect to c.

As $\tau < \sigma_0$, we find, indicating by accents the σ's for (7), that, if $\phi_1^{(i)}(c_1) \neq 0$, then

$$\sigma_i' = \sigma_0 - i\rho_1, \tag{13}$$

but that, if $\phi_1^{(i)}(c_1) = 0$,

$$\sigma_i' < \sigma_0 - i\rho_1. \tag{14}$$

Thus, if c_1 is a root of (6) of multiplicity s_1, we have (14) for $i < s_1$, in particular, for $i = 0$, and (13) for $i = s_1$.

We have

$$\frac{\sigma_0' - \sigma_i'}{i} = \frac{\sigma_0' - \sigma_0}{i} + \frac{\sigma_0 - \sigma_i'}{i}. \tag{15}$$

From (13) and (14), we see that $(\sigma_0 - \sigma_i')/i$ is a minimum, namely ρ_1, for $i = s_1$. As (14) holds for $i = 0$, we have $\sigma_0' < \sigma_0$, so that $(\sigma_0' - \sigma_0)/i$ is greater for $i > s_1$ than for $i = s_1$. Hence the greatest value of i for which the first member of (15) is a minimum cannot exceed s_1. This proves that $g_2 \leqq s_1$.

Furthermore, for $i = s_1$, the first member of (15) is ρ_1 decreased by $(\sigma_0 - \sigma_0')/s_1$, so that ρ_2, the minimum of the first member of (15), is less than ρ_1. Similarly $\rho_{p+1} < \rho_p$ for every p.

6. We are going to prove (c) of §4, and to complete the proof of (a).

With every positive integer p, we have associated a c_p, ρ_p, g_p, and s_p. For every p,

$$g_{p+1} \leqq s_p \leqq g_p \leqq n.$$

Thus there exists a q such that

$$g_p = s_p = g_q$$

for $p \geqq q$. Let g_q be represented by s. For $p \geqq q$, the s roots of $\phi_p(c) = 0$ are all equal. Then

$$\phi_p(c) = b_p(c - c_p)^s$$

with b_p a constant.

We consider the relation, analogous to (10),

$$f_{p-1}(ce^{\rho_p z}) = \phi_p(c)e^{\sigma_0^{(p-1)}z} + \psi_p(c, z)e^{\tau^{(p-1)}z},$$

and put in it $c = w^{(p-1)}e^{-\rho_p z}$. We find, for $p \geqq q$,

$$f_{p-1}(w^{(p-1)}) = b_p(w^{(p-1)}e^{-\rho_p z} - c_p)^s e^{\sigma_0^{(p-1)}z} + \psi_p(w^{(p-1)}e^{-\rho_p z}, z)e^{\tau^{(p-1)}z}.$$

Differentiating $s-1$ times with respect to $w^{(p-1)}$, we find

$$f_{p-1}^{(s-1)}(w^{(p-1)}) = s!b_p e^{(\sigma_0^{(p-1)} - (s-1)\rho_p)z}(w^{(p-1)}e^{-\rho_p z} - c_p) + \psi_p^{(s-1)}(w^{(p-1)}e^{-\rho_p z}, z)e^{(\tau^{(p-1)} - (s-1)\rho_p)z}. \tag{16}$$

The expression $\psi_p^{(s-1)}$ in (16) is the result of replacing c by $w^{(p-1)}e^{-\rho_p z}$ in the $(s-1)$th derivative with respect to c of $\psi_p(c, z)$.

We shall now find a function $w^{(p-1)}$ of z such that

$$f_{p-1}^{(s-1)}(w^{(p-1)}) = 0.$$

We must have, by (16),

$$w^{(p-1)} = c_p e^{\rho_p z} - \frac{1}{s!b_p}\psi_p^{(s-1)}(w^{(p-1)}e^{-\rho_p z}, z)e^{(\tau^{(p-1)} - \sigma_0^{(p-1)})z + \rho_p z}. \tag{17}$$

We put

$$w^{(p-1)} = (u + c_p)e^{\rho_p z}.$$

Then (17) becomes

$$u = -\frac{1}{s!b_p}\psi_p^{(s-1)}(u + c_p, z)e^{(\tau^{(p-1)} - \sigma_0^{(p-1)})z}. \tag{18}$$

Since $\psi_p^{(s-1)}(u+c_p, z)$ is a polynomial in u whose coefficients are absolutely convergent Dirichlet series with non-positive exponents, and since $\tau^{(p-1)} < \sigma_0^{(p-1)}$, we can write (18)

$$u = Q_0 + Q_1 u + \cdots + Q_n u^n, \tag{19}$$

where the Q's are absolutely convergent Dirichlet series, with *negative* exponents, of the type (4). When $x \to +\infty$, every Q approaches 0.

Now if the Q's are regarded as independent variables, (19) has a unique solution for u which is analytic when every Q is small, and which equals 0 when the Q's are all zero. Let this solution be

$$u = F(Q_0, \cdots, Q_n), \tag{20}$$

where F is a series of positive powers in the Q's. If the Q's are replaced by the Dirichlet series which correspond to them, u becomes a Dirichlet series with a half-plane of absolute convergence,

$$u = d_1 e^{\tau_1 z} + d_2 e^{\tau_2 z} + \cdots, \tag{21}$$

where $0 > \tau_1 > \tau_2 > \cdots$, $\lim \tau_m = -\infty$.

Then the equation $f_{p-1}^{(s-1)}(w^{(p-1)}) = 0$ is satisfied by the Dirichlet series

$$w^{(p-1)} = c_p e^{\rho_p z} + d_1 e^{(\tau_1 + \rho_p) z} + \cdots,$$

which has a half-plane of absolute convergence.

Now

$$f_{p-1}^{(s-1)}(w^{(p-1)}) = f^{(s-1)}(c_1 e^{\rho_1 z} + \cdots + c_{p-1} e^{\rho_{p-1} z} + w^{(p-1)}).$$

Hence the equation

$$f^{(s-1)}(w) = 0 \tag{22}$$

is satisfied by

$$w = c_1 e^{\rho_1 z} + \cdots + c_p e^{\rho_p z} + d_1 e^{(\tau_1 + \rho_p) z} + \cdots. \tag{23}$$

For every p, there is a solution of (22) of type (23). Of course, the τ's and d's, in (21), depended, as far as we knew, on p, but as (22) has at most n solutions, and as two distinct series (23) cannot represent the same function, there can be at most n series (23) for the various values of p. One of these series must be the series (8). Hence $\lim \rho_p = -\infty$, and (8) has a half-plane of absolute convergence. This proves (a) and (c).

7. To prove (d), it will suffice to show that (8) satisfies (2) *formally*. This will follow if we can show that $\sigma_0^{(p)} \to -\infty$ as $p \to \infty$. As f_p is obtained from f by replacing w by $w^{(p)}$ plus a segment of (8), it is obvious that an M exists such that $\sigma_i^{(p)} < M$ for every i and p. Now, for $p \geqq q$, by (13),

$$M > \sigma_s^{(p+1)} = \sigma_0^{(p)} - s\rho_{p+1}$$

so that $\sigma_0^{(p)} < M + s\rho_{p+1}$, and $\sigma_0^{(p)} \to -\infty$, because $\rho_{p+1} \to -\infty$.

II. Dirichlet series with complex exponents

8. In what follows, series represented by small letters (as $a(z)$) will be series

$$v_1 e^{\mu_1 iz} + \cdots + v_p e^{\mu_p iz} + \cdots , \tag{24}$$

where the μ's are real and either increase with their subscripts, tending toward $+\infty$, or else decrease as their subscripts increase, tending toward $-\infty$. The v's are constants.

Series represented by large letters (as $P(z)$) will be of the form

$$a_1(z)e^{\lambda_1 z} + \cdots + a_p(z)e^{\lambda_p z} + \cdots , \tag{25}$$

where the λ's are real and decreasing, with $-\infty$ as their limit. The a's are of type (24), and while the μ's in any two of them need not be the same, the μ's must either increase with their subscripts in all a's, or else decrease in all a's.

9. We shall use, in what follows, sectors of the complex plane bounded by two half-lines which emanate from a point. The boundary of a sector will be understood to belong to the sector.

10. If $a(z)$ is of type (24), we shall use $\bar{a}(z)$ to represent the sum of the moduli of the terms of $a(z)$.

We shall say that (25) converges *normally* in a sector, if every $\bar{a}(z)$, and also

$$\sum_{p=1}^{\infty} \bar{a}_p(z) \left| e^{\lambda_p z} \right| , \tag{26}$$

converges in the sector, and if, for every $h>0$, a positive integer q exists such that

$$\sum_{p=q+1}^{\infty} \bar{a}_p(z) \left| e^{\lambda_p z} \right| < e^{-h|z|} \tag{27}$$

for z in the sector.

11. We prove that *if $P(z)$ is normally convergent in a sector S, then $P(z)$ is uniformly convergent in S.*

It suffices to prove that $P(z)$ is uniformly convergent in any finite triangular part of S. The normal convergence of $P(z)$ will then guarantee that the convergence is uniform throughout S.

The uniformity of the convergence in any triangle is essentially a result of Hille.* Let the vertices of the triangle be z_1, z_2, z_3. Then any z within or on the boundary of the triangle has a representation

* Hille, p. 263.

$$z = \gamma_1 z_1 + \gamma_2 z_2 + \gamma_3 z_3,$$

where the γ's are non-negative numbers whose sum is unity. Then, for any β,

$$\left| e^{\beta z} \right| = \left| (e^{\beta z_1})^{\gamma_1} (e^{\beta z_2})^{\gamma_2} (e^{\beta z_3})^{\gamma_3} \right| \leqq \max \left| e^{\beta z_i} \right|, \quad i = 1,2,3.$$

It is thus clear that the remainder of (25) at any point in the triangle has a modulus not greater than the sum of the corresponding remainders of (26) at the points z_1, z_2, z_3. This shows that the convergence is uniform.

If (26) is convergent in S, we know that $P(z)$ converges uniformly in any finite part of S. Thus to prove that $P(z)$ is normally convergent in S, it suffices to show that (26) converges in S and that for every $h>0$, an $r>0$ can be found such that some remainder of (26) is less than $e^{-h|z|}$ for $|z|>r$.

12. Let S be a sector with one side extending in the direction of the positive real axis. Let

$$P(z) = \sum a_p(z) e^{\lambda_p z}, \qquad Q(z) = \sum b_p(z) e^{\lambda_p z},$$

where each pair a_p and b_p have the same μ's, be normally convergent in S. Furthermore, suppose that $P(z) = Q(z)$ in S. We shall prove that $a_p(z) \equiv b_p(z)$ for every p.

It will suffice to prove that $a_1(z) \equiv b_1(z)$.

Consider any half-line in S which is directed like the positive real axis. We have, in S,

$$a_1(z) + \sum_{p=2}^{\infty} a_p(z) e^{(\lambda_p - \lambda_1)z} = b_1(z) + \sum_{p=2}^{\infty} b_p(z) e^{(\lambda_p - \lambda_1)z}. \tag{28}$$

Because the two series in (28) are normally convergent, the remainder after a large number of terms in either series approaches zero as z goes to ∞ on the half-line taken above. But as every $\lambda_p - \lambda_1$ is negative for $p>1$ and as every a_p and b_p is bounded on the half-line, the sum of any finite number of terms of either series approaches zero as z increases on the half-line. Hence, for such an increase of z, both series approach 0. Now $a_1(z)$ and $b_1(z)$, functions defined by absolutely convergent Dirichlet series, are almost periodic, with real quasi-periods. Thus there is a sequence of positive numbers $h_1, \cdots, h_p, \cdots$, tending toward $+\infty$, such that, for any z on the half-line,

$$a_1(z + h_p) \to a_1(z), \qquad b_1(z + h_p) \to b_1(z)$$

as $p \to \infty$.* This, by (28), means that $a_1(z) = b_1(z)$ for every z on the half-line. Hence $a_1(z)$ is identical with $b_1(z)$.

---

* Bohr, Acta Mathematica, vol. **45** (1925), p. 37.

13. We consider two series of type (25)

$$P = \sum a_p(z)e^{\lambda_p z}, \qquad Q = \sum b_p(z)e^{\lambda_p z},$$

where a_p and b_p have the same μ's, which tend toward $+\infty$. We assume that P and Q are both normally convergent in a sector S, which has one side extending in the direction of the positive real axis, and whose second side is inclined at a positive angle less than π to the first one. It is understood that S is generated by a positive rotation from its first side to its second.*

The product PQ can be written as a series of type (25), $\sum c_p(z)e^{\lambda_p' z}$ where each λ' is a sum of two λ's and where each c is a bilinear combination of a's and b's. Furthermore, every $\bar{c}_p(z)$ and also $\sum \bar{c}_p(z)\,|e^{\lambda_p' z}|$ will be convergent in S.

We are going to prove that PQ is normally convergent in S.

It is easy to see that the product of P by an $e^{\lambda z}$ or by an absolutely convergent series of type (24), with increasing μ's, is normally convergent in S. (Cf. final remarks of §11.) Hence the product of P by the sum of a finite number of terms of Q is normally convergent in S.

Let $h>0$ be assigned arbitrarily, and let q be taken so that

$$\sum_{p=q+1}^{\infty} \bar{a}_p(z)\,|e^{\lambda_p z}| < e^{-h|z|}, \qquad \sum_{p=q+1}^{\infty} \bar{b}_p(z)\,|e^{\lambda_p z}| < e^{-h|z|}$$

in S. Let, furthermore,

$$P = a_1(z)e^{\lambda_1 z} + \cdots + a_q(z)e^{\lambda_q z} + P_q(z),$$
$$Q = b_1(z)e^{\lambda_1 z} + \cdots + b_q(z)e^{\lambda_q z} + Q_q(z).$$

Now PQ is the sum of an infinite number of infinite series of the form $c(z)e^{\lambda' z}$. We shall consider PQ as a double series of terms $ve^{(\mu i+\lambda')z}$. Those terms of the double series which come from multiplying terms of P_q and Q_q have less than $e^{-2h|z|}$ for the sum of their moduli. From this, and from the fact that the product of P_q, (Q_q), by the sum of the first q terms of Q, (P), is normally convergent in S, it follows that PQ is normally convergent in S.

Of course, we can derive a similar result for a's and b's with decreasing μ's, provided that the second side of S makes a negative angle less than π with the first side.

14. Let S be a sector which has one side extending in the direction of the positive real axis, and whose second side makes a positive angle less than π with the first one.† (See Fig. 1.)

* This will always be understood when we say that the opening of a sector is a positive angle.

† The assumption that the opening is less than π has no bearing on our final results, but it contributes simplicity to our discussion.

Let $Q_1, \cdots, Q_n$ be n series of type (25), with *negative* λ's and increasing μ's, all normally convergent in S.

Let $u=F(u_1, \cdots, u_n)$ be a function analytic in all of its arguments at $(0, \cdots, 0)$.

Let $\overline{Q}_i$ be the series (26) for Q_i. Consider any term $\bar{a}(z)\,|e^{\lambda z}|$ of $\overline{Q}_i$. Because $\lambda<0$, there is an $h>0$ such that $\bar{a}(z)\,|e^{\lambda z}|<e^{-h|z|}$ for z sufficiently large, in a sector whose first side is the horizontal side of S and whose second side makes a sufficiently small positive angle with the first one. For, if the opening of the new sector is made sufficiently small, $\bar{a}(z)$, if it becomes infinite at all, as $z\to\infty$, will become infinite like $e^{k|z|}$ with k small.

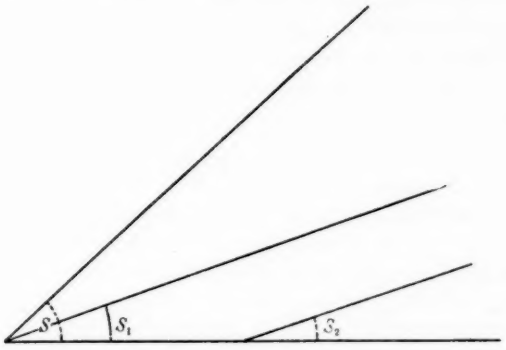


FIG. 1

From this fact, and from the normal convergence of Q_i, we see that there is an $h>0$ and a sector S_1, contained in S and with one side coincident with the horizontal side of S, such that, for every i, $\overline{Q}_i<e^{-h|z|}$ for z large and in S_1. In what follows, we suppose an $h>0$ and an S_1 to have been found, and we hold them fixed.

Let $\overline{F}$, a majorant for $F(u_1, \cdots, u_n)$, converge for $|u_i|<\delta$, where $\delta>0\,(i=1, \cdots, n)$. Let S_2 be a sector contained in S_1, with vertex on the horizontal side of S_1, with one side horizontal and with its second side parallel to the second side of S_1, such that $\overline{Q}_i<\delta$ in S_2, for every i.

When each u_i is replaced by Q_i in $u=F(u_1, \cdots, u_n)$, u will become a series $U(z)$ of type (25), absolutely convergent (considered as a double series) in S_2.

We say that $U(z)$ is normally convergent in S_2.

For t any integer, arbitrarily large, let

$$F(u_1, \cdots, u_n) = A(u_1, \cdots, u_n) + B(u_1, \cdots, u_n),$$

where A consists of those terms of F which are of degree not exceeding t, and where B consists of the remaining terms of F. Let $\overline{B}$ be the majorant of B. Then there is an $\eta>0$, depending on t, such that, for $|u_i|<\eta$,

$$|\overline{B}| \leqq \max(|u_1|^t, \cdots, |u_n|^t). \tag{29}$$

When each u_i is replaced in A by Q_i, we get, according to §13, a series (25) normally convergent in S, hence in S_2. But, by (29), B will yield a double series the sum of the moduli of the terms of which is less than $e^{-\lambda t|z|}$ for z large in S_1. Furthermore, t may be taken arbitrarily large. These facts and the final remarks of §11, show that $U(z)$ converges normally in any sector contained in S_1 in which $U(z)$ converges (as a double series) absolutely. The sector S_2 is such a sector contained in S_1.

A similar result holds for the case where the μ's are decreasing and the second side of S makes a negative angle with the first.

15. We consider a sector AMB (Fig. 2) consisting of those points z, for which

$$|\operatorname{amp}(z-M)| \leqq \theta,$$

where $0<\theta<\pi$, the point M being on the real axis. Let O_1 be on MA, O_2 be conjugate to O_1. Let O_1C_1 and O_2C_2 be horizontal half-lines having the direction of the positive real axis.

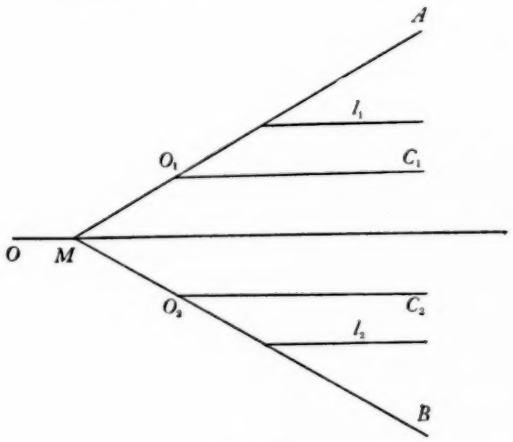


FIG. 2

Let $\phi(z)$ be a function analytic in the sector AMB. Let $\phi(z)$ have in AO_1C_1 a normally convergent representation

$$a_1(z)e^{\lambda_1 z} + \cdots + a_p(z)e^{\lambda_p z} + \cdots \tag{30}$$

with increasing μ's, and, in BO_2C_2, a normally convergent representation

$$b_1(z)e^{\lambda_1 z} + \cdots + b_p(z)e^{\lambda_p z} + \cdots \tag{31}$$

with the same* λ's which appear in (30) but with decreasing μ's.

Furthermore suppose that a $k>0$ exists such that $|\phi(z)| < e^{k|z|}$ in AMB.

We say that

(a) *For every p, $a_p(z)$ and $b_p(z)$ are exponential polynomials, and $a_p(z) \equiv b_p(z)$.*

(b) *The identical series* (30) *and* (31) *converge normally to $\phi(z)$ in a sector interior to AMB, with vertex on the real axis, and with sides parallel to those of AMB.*

In proving (a), we may limit ourselves to the case of $p=1$.

Let l_1 and l_2 be horizontal half-lines respectively interior to AO_1C_1 and BO_2C_2 and at equal distances from the real axis. We take any half-strip between l_1 and l_2, that is, a region contained in AMB, limited by l_1, l_2 and a segment perpendicular to l_1 and l_2.

Consider the function $\psi(z)e^{-\lambda_1 z}\phi(z)$. By (30), (31), it is bounded on l_1 and l_2. Hence as $|\psi(z)| < e^{(k+|\lambda_1|)|z|}$, in the half-strip between l_1 and l_2, $\psi(z)$, according to the Phragmen-Lindelöf theorem, must be bounded in the half-strip.

The functions $a_1(z)$ and $b_1(z)$ are almost periodic, with real quasi-periods, on O_1C_1 and O_2C_2 respectively. We consider a sequence of positive numbers $h_1, \cdots, h_p, \cdots$, increasing to infinity, such that for z on O_1C_1, $a_1(z+h_p) \to a_1(z)$ as $p\to\infty$, and such that, for z on O_2C_2, $b_1(z+h_p)\to b_1(z)$ as $p\to\infty$.

Then $\psi(z+h_p)\to a_1(z)$ on O_1C_1 and $\psi(z+h_p)\to b_1(z)$ on O_2C_2 as $p\to\infty$.

Consider any rectangle cut from the half-strip between l_1 and l_2 by vertical lines. In such a rectangle, the functions $\psi(z+h_p)$ have a common upper bound for their moduli. The sequence $\psi(z+h_p)$ converges on two horizontal lines inside of the rectangle. By the Vitali-Porter theorem, the sequence converges uniformly in any area interior to the rectangle. The analytic limit approached by $\psi(z+h_p)$ equals $a_1(z)$ on O_1C_1 and $b_1(z)$ on O_2C_2. Hence $a_1(z)$ and $b_1(z)$ are continuations of each other. That is, there is a function $c(z)$, analytic in a right half-plane T, which coincides with $a_1(z)$ above O_1C_1, and with $b_1(z)$ below O_2C_2. Furthermore, $c(z)$ is bounded in any horizontal half-strip of T.

It is now easy to see that $a_1(z)$ contains only a finite number of μ's. For if $v_m e^{\mu_m i z}$ is a term of $a_1(z)$, then, ω being a positive variable,

$$v_m = \lim_{\omega\to\infty} \frac{1}{\omega} \int_{O_1}^{O_1+\omega} e^{-\mu_m i z} c(z)\,dz, \tag{32}$$

* The assumption that the λ's are the same is not a restrictive one

the integration being performed along O_1C_1. Now, as $e^{-\mu_m iz}c(z)$ is bounded in any horizontal half-strip of T, we have also

$$v_m = \lim_{\omega\to\infty} \frac{1}{\omega} \int_{O_2}^{O_2+\omega} e^{-\mu_m iz} c(z)dz. \tag{33}$$

For the difference between the integrals in (32) and (33) is

$$\int_{O_1}^{O_2} - \int_{O_1+\omega}^{O_2+\omega},$$

which is bounded for all ω's.

Imagine that μ_m exceeds the first μ in $b_1(z)$. As the μ's in $b_1(z)$ decrease, μ_m will be distinct from every μ in $b_1(z)$. By (33) we will have $v_m=0$. Hence $a_1(z)$ has only a finite number of μ's and is an exponential polynomial. The series $b_1(z)$ is an exponential polynomial identical with $a_1(z)$.

To prove (b), we consider any term $w(z)=ve^{(\mu i+\lambda)z}$ of (30) regarded as a double series. Let z be any point between O_1C_1 and O_2C_2, and z_1, z_2 two points of the same abscissa as z, but respectively above O_1C_1 and below O_2C_2. Then*

$$z=(1-k)z_1+kz_2,$$

where $0<k<1$. Hence

$$|w(z)| = |w(z_1)|^{1-k}|w(z_2)|^k \leqq \max(|w(z_1)|, |w(z_2)|),$$

and this is easily seen to imply (b).

16. Let

$$P(z) = \sum a_p(z)e^{\lambda_p z}, \qquad Q(z) = \sum b_p(z)e^{\lambda_p z}, \tag{34}$$

with every $a(z)$ and every $b(z)$ an exponential polynomial, be normally convergent in a sector S, of opening less than 2π, which is symmetrical with respect to, and contains an infinite portion of, the positive real axis.

Suppose that P/Q is analytic in S.

We shall prove that there is a sector T, contained in S, symmetric with respect to, and containing an infinite portion of, the positive real axis, in which P/Q admits a normally convergent expansion $\sum c_p(z)e^{\lambda' z}$, with every $c(z)$ an exponential polynomial.

We shall show first that in some sector S_1, contained in S, with one side in the direction of the positive real axis, and generated by a positive rotation less than π, the quotient P/Q has a normally convergent expansion (25) with increasing μ's.

Let t be the smallest value of p for which $b_p(z)\not\equiv 0$. Then $Q(z)=b_t(z)e^{\lambda_t z}$

* Cf. Hille, p. 263.

$\cdot (1+R(z))$, where $R(z)$ has a normally convergent expansion (25), with increasing μ's, and with *negative* λ's, in a sector U like the S_1 just described. By §14, $1/(1+R(z))$ has a normally convergent expansion (25) in a sector V with one side horizontal, interior to U. Then $1/Q(z)$ has a normally convergent expansion (25) in a sector S_1, described as above, interior to V. By §13, P/Q has an expansion (25) with increasing μ's, normally convergent in S_1.

Similarly, in a sector S_2, generated by a negative rotation, P/Q has a normally convergent expansion (25) with decreasing μ's.

We are going to show that the expansions in S_1 and S_2 are identical, and that the $a(z)$'s in them are exponential polynomials. This, according to the proof of (b) of §15, will give us our result.

Let $a_q(z)$ and $b_t(z)$ be, respectively, the first $a(z)$ and $b(z)$ which are not identically zero. We shall show that $a_q(z)/b_t(z)$ is an exponential polynomial with pure imaginary exponents.

It is no specialization to assume that $q=t=1$ and that $\lambda_1=0$. These circumstances may be brought about by multiplying P and Q by exponentials and using a new set of λ's.

There exists a horizontal strip L which contains all of the zeros of $b_1(z)$ in its interior. Let B be the difference between the greatest and least μ's in $b_1(z)$. The number of zeros of $b_1(z)$ whose abscissas lie between u and v $(u<v)$ is equal to*

$$\frac{B(v-u)}{2\pi}+O(1).$$

We may obviously take L so broad that $b_1(z)$ is bounded away from zero along the sides of L. We suppose this to be done.

Because $b_1(z)$ is almost periodic, we can break L up into an infinite number of rectangles, whose lengths have a common upper bound, such that for some $h>0$, $|b_1(z)|>h$ along all of the vertical sides of the rectangles.

Now $Q(z)=b_1(z)+R(z)$, where $R(z)$ has negative λ's. $R(z)$ approaches zero as z goes towards ∞ in L, in the right-hand direction.

Consequently, in any of the rectangles mentioned above, situated at a great distance to the right, $Q(z)$ has the same number of zeros as $b_1(z)$. It follows easily that the number of zeros of $Q(z)$, in the region common to L and to S, with abscissas between u and v, is equal to

$$\frac{B(v-u)}{2\pi}+O(1). \tag{35}$$

* Tamarkin, p. 24, Pólya, Schwengler, Wilder, p. 420, Ritt (1).

Suppose that a_1/b_1 is not an exponential polynomial. Then it is possible to find an exponential polynomial $g(z)$, with pure imaginary exponents, the difference between whose greatest and least μ's is less than B, above, such that

$$a_1(z) = d(z)b_1(z) + g(z),$$

with $d(z)$ an exponential polynomial having pure imaginary exponents.*

We consider the function $P(z)-d(z)Q(z)$, which, because P vanishes wherever Q does, in S, also vanishes where Q does, in S. We have

$$P(z) - d(z)Q(z) = g(z) + K(z),$$

where $K(z)$ is a series of type (25), with *negative* λ's. This means, by the discussion leading up to (35), that $P(z)-d(z)Q(z)$ has fewer zeros than $Q(z)$ in a long horizontal rectangle. We have here a contradiction which shows that $a_1(z)/b_1(z)$ is an exponential polynomial.

Thus $a_q(z)/b_t(z)$ is, as we stated, an exponential polynomial, $c_1(z)$.

Now the first coefficient in the expansion of $1/Q$ in S_1 is $1/b_t(z)$ expanded with increasing μ's. Hence the first coefficient in the expansion (25) of P/Q is $a_q(z)/b_t(z)$, which is $c_1(z)$. The same is true of the expansion of P/Q in S_2.

Let the first term in the expansion of P/Q, either in S_1 or in S_2, be $c_1(z)e^{\lambda_1' z}$.

The second term in the expansion of P/Q, say in S_1, is the first term in the expansion of

$$\frac{P - c_1(z)e^{\lambda_1' z}Q}{Q}. \tag{36}$$

Since (36) is analytic in S, we know that the first coefficient in its expansion is an exponential polynomial.

Continuing in this way, we find that the developments of P/Q in S_1 and in S_2 are identical, and that the coefficients in them are exponential polynomials. We thus have the result stated at the beginning of this article.

17. Let OX (Fig. 3) be the positive real axis, and OA a ray of positive inclination less than π. Let S be any sector contained in XOA, with sides parallel to those of XOA. Let

$$P(z) = a_1(z)e^{\lambda_1 z} + \cdots + a_p(z)e^{\lambda_p z} + \cdots, \tag{37}$$

with increasing μ's, be normally convergent in S.

* Ritt, (1).

Let ϕ be a positive angle less than angle XOA. Suppose that, in (37), we replace z by $e^{i\phi}z$.

Consider any term

$$ve^{\mu iz}e^{\lambda z} \tag{38}$$

of (37) regarded as a double series. This term goes over into

$$ve^{(\mu\cos\phi+\lambda\sin\phi)iz}e^{(\lambda\cos\phi-\mu\sin\phi)z}. \tag{39}$$

We shall prove that, given any real h, there are only a finite number of pairs μ, λ such that $\lambda\cos\phi-\mu\sin\phi>h$.

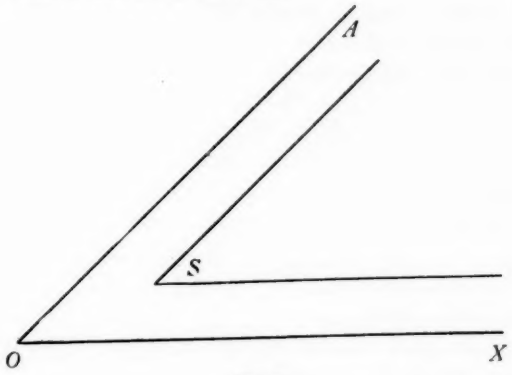


FIG. 3

We shall thus know that $P(e^{i\phi}z)$ can be written in the form

$$b_1(z)e^{\lambda_1' z}+\cdots+b_p(z)e^{\lambda_p' z}+\cdots, \tag{40}$$

with λ's decreasing toward $-\infty$, and with every b an exponential polynomial. Furthermore, (40) will converge absolutely to $P(e^{i\phi}z)$ in the sector obtained by rotating S about the origin through an angle $-\phi$.

For the proof, consider any term of (37). There can be in it only a finite number of terms like (38) for which $\lambda\cos\phi-\mu\sin\phi>h$, for any h, since $\sin\phi>0$ and the μ's in any term of (37) increase to $+\infty$. Again, because of the normal convergence of (37), a term (38) coming from a large p of (37) must approach zero along the line amp $z=\phi$ like an e^{-kz} with k very large and positive. A similar fact is therefore true of (39) along the positive real axis. Hence $\lambda\cos\phi-\mu\sin\phi$ must be large and negative when p is large. This proves our statement.

Let T be the sector into which S goes, after a rotation about the origin through an angle $-\phi$. Let U be a sector whose vertex is that of T, whose

first side has the inclination $-\phi+\epsilon$, where ϵ is arbitrarily small and positive, and whose second side is the second side of T. That is, U is found from T by cutting away a small sector at the lower edge of T.

We say that, in U, (40) converges *normally* to $P(e^{i\phi}z)$.

We take any $a_p(z)$ of (37), and consider it as a Dirichlet series. For any sector of opening less than π, symmetrical with respect to the positive imaginary axis, and for any $h>0$, the remainder after a large number of terms in $\bar{a}_p(z)$ is less than $e^{-h|z|}$.

Hence, in the sector U, the sum of the moduli of those terms (39) which come from a particular $a_p(z)e^{\lambda_p z}$, and whose μ's are sufficiently large, will be less than $e^{-h|z|}$, where h is large at pleasure.

On the other hand, because of the normal convergence of (37) in S, we can fix an m such that the terms (39) which come from the remainder after m terms in (37) have, in T, a sum for their moduli less than $e^{-h|z|}$.

For p sufficiently great, in (40), the terms (39) in $b_p(z)e^{\lambda'_p z}$ come either from a term beyond the mth in (37) or else from terms (38) of large μ in one of the first m terms of (37).

Thus, for p large,

$$\sum_{i=p+1}^{\infty} \bar{b}_i(z) \mid e^{\lambda'_i z} \mid < (m+1)e^{-h|z|}. \tag{41}$$

For any $k<h$, the second member of (41) is, for $|z|$ large, less than $e^{-k|z|}$. This, by the final remarks of §11, proves our statement.

III. Equations whose coefficients have complex exponents

18. We consider an equation

$$f(w) = P_n(z)w^n + \cdots + P_1(z)w + P_0(z) = 0, \tag{42}$$

where every $P(z)$ is of type (25), with increasing μ's, normally convergent in a sector S whose first side has the direction of the positive real axis, the sector being generated by a positive rotation less than π, starting from the first side. We assume that $P_n(z)\not\equiv 0$.

We shall prove that the n solutions of (42) are all analytic in a sector U interior to S, the sector U having one side in the direction of the positive real axis and being generated by a positive rotation starting from its first side. In U, the n solutions will each have a normally convergent representation (25), with increasing μ's. Each λ in the representations of the solutions will be given by an expression like the second member of (4), in terms of the λ's in the P's. Each μ in the representations of the solutions will be similarly expressed in terms of the μ's in the P's.

For the case in which the P's have decreasing μ's and are normally convergent in a sector generated by a negative rotation, similar results hold, the solutions having representations (25) with decreasing μ's.

We assume that $P_0(z) \not\equiv 0$. For every i, we define σ_i as in §3, using series $a(z)$ instead of constants a. Similarly, we define g_1 and ρ_1.

We consider the equation (6), where now each $\alpha^{(1)}$ is a series (24) with increasing μ's. The g_1 solutions of (6) will be analytic in some upper half-plane, and will have absolutely convergent expansions (24) with μ's found from the μ's in the $\alpha^{(1)}$'s by expressions like the second member of (4), with r fixed. Let c_1 be any solution of (6).

We put, in (42), $w = c_1 e^{\rho_1 z} + w'$. Then w' is determined by (7), where, now, each P' is of type (25), normally convergent in S_1, the sector common to S and the half-plane of absolute convergence of c_1.

As in §3, we build a series

$$c_1(z)e^{\rho_1 z} + c_2(z)e^{\rho_2 z} + \cdots + c_p(z)e^{\rho_p z} + \cdots .$$

We continue as in §5. In (10), $\psi_1(c, z)$ is a polynomial in c, with coefficients of type (25), which have non-positive λ's and are normally convergent in S. In (12), $\psi_1^{(i)}(c_1, z)$ is normally convergent in S_1 and has non-positive λ's. If s_1 is defined as in §5, we find again that $g_2 \leqq s_1$. Also, it is seen that $\rho_2 < \rho_1$.

We arrive finally at (19) of §6, where now each Q is of type (25), with *negative* λ's, normally convergent in a sector S_p which has one side in the direction of the positive real axis, and whose second side is the second side of S.

Let $\overline{Q}_i$ be defined as in §14. We see, as in §14, that in some sector T whose first side is the horizontal side of S_p and whose second side makes a sufficiently small positive angle with the first side, every $\overline{Q}_i \to 0$ as $|z| \to \infty$.

Hence we can replace every Q_i in (20) by its expansion (25), and get an expression (25) for u, normally convergent in a sector U with vertex on the horizontal side of T, the first side of U being horizontal, and the second side of U being parallel to the second side of T.

The function u thus obtained is a solution of (18). It leads, as in §§6, 7, to a solution of (42) of type (25).

IV. Applications to algebraic combinations of exponentials

19. We consider now an equation

$$P_n w^n + P_{n-1} w^{n-1} + \cdots + P_0 = 0, \tag{43}$$

with every P an exponential polynomial and $P_n \not\equiv 0$.

Evidently a function w, defined by (43), has, in the finite part of the

plane, only isolated singularities, which are poles or branch points. When $P_n \equiv 1$, there are no poles.

If a sector T is contained in a sector S, and if the sides of T have the same directions as those of S, we shall call T a *subsector* of S. The vertex of T may be on the boundary of S.

Consider a series

$$v_1 e^{\sigma_1 z} + \cdots + v_p e^{\sigma_p z} + \cdots \tag{44}$$

with constant v's, and with real or complex σ's such that, as $p \to \infty$, $|\sigma_p| \to \infty$.

We shall say that (44) is *normally* convergent in a sector, if it is absolutely convergent in the sector, and if, for every $h > 0$, a q exists such that

$$\sum_{p=q+1}^{\infty} |v_p e^{\sigma_p z}| < e^{-h|z|}$$

throughout the sector. We see, as in §11, that (44) converges uniformly in any sector in which it converges normally.

Let S be a sector of opening less than 2π, and suppose that some branch $w = f(z)$ of a function defined by (43) is analytic in a sector which has the same vertex as S, and which contains in its interior every point of S other than the vertex.

We are going to show that, in some subsector of S, $w = f(z)$ has a normally convergent development of type (44). The real part of each σ is expressed in terms of the real parts of the exponents of the P's by a relation like (4), and the imaginary parts of the σ's are expressed analogously.

Let the angles between the sides of S and the positive real axis be ϕ_1 and ϕ_2, where

$$0 \leqq \phi_1 < \phi_2 \leqq 2\pi, \quad \phi_2 - \phi_1 < 2\pi.$$

Let ψ be such that $\phi_1 \leqq \psi \leqq \phi_2$.

We shall show that $f(e^{i\psi}z)$ has an expansion (25), with a's which are exponential polynomials, normally convergent in some sector symmetrical to, and containing an infinite portion of, the positive real axis.

When P_n is a constant, this follows quickly from §15, for $f(e^{i\psi}z)$ satisfies all of the hypotheses of that article when P_n is a constant.

When P_n is not a constant, we make the substitution $P_n w = y$, whereupon (43) goes over into an equation in y with unity for the coefficient of y^n. Thus $P_n(e^{i\psi}z)f(e^{i\psi}z)$ satisfies the hypotheses of §15. Also the pair of functions $P_n(e^{i\psi}z)f(e^{i\psi}z)$ and $P_n(e^{i\psi}z)$ satisfy the hypotheses of §16, so that we have our result for $f(e^{i\psi}z)$.

Let the expansion of $f(e^{i\psi}z)$ be normally convergent in a sector symmetric

to the positive real axis, of opening $2\delta_\psi < 2\pi$. By §17, for $|\eta| < \delta_\psi$, we can get the expansion (25) of $f(e^{i(\psi+\eta)}z)$ from that of $f(e^{i\psi}z)$ by replacing z by $e^{i\eta}z$. (Evidently the result of §17 holds, for decreasing μ's, when OA has a negative inclination greater than $-\pi$.)

Consider the intervals $(\psi-\delta_\psi, \psi+\delta_\psi)$ and pick out a finite number of them whose interiors cover the entire interval (ϕ_1, ϕ_2).

We see now that, for every ψ, the expansion of $f(e^{i\psi}z)$ can be found from that of $f(e^{i\phi_1}z)$ by replacing z by $e^{i(\psi-\phi_1)}z$; for we can pass from ϕ_1 to ψ across the small intervals, in a finite number of steps.

We now consider the expansion (25) of $f(e^{i\phi_1}z)$ as a double series and order its terms so that the moduli of the exponents are non-decreasing. We have thus

$$f(e^{i\phi_1}z) = v_1e^{\tau_1 z} + \cdots + v_pe^{\tau_p z} + \cdots, \tag{45}$$

where the τ's are real or complex numbers such that $|\tau_p| \to \infty$ as $p \to \infty$.

We have from (45)

$$f(z) = v_1e^{\sigma_1 z} + \cdots + v_pe^{\sigma_p z} + \cdots, \tag{46}$$

where $\sigma_p = e^{-i\phi_1}\tau_p$.

It is now easy to see that the second member of (46) converges normally to $f(z)$ in a subsector of S.

For the expansion (25) of $f(e^{i\psi}z)$, $\phi_1 \leqq \psi \leqq \phi_2$, is obtained from (46) by replacing z by $e^{i\psi}z$. Now as the expansion (25) of $f(e^{i\psi}z)$ has a's which are exponential polynomials, the distant terms of (46) give distant terms in the expansion of $f(e^{i\psi}z)$. Hence, for every ψ, (46) is normally convergent in a sector symmetrical with respect to the line amp $z=\psi$. The Heine-Borel theorem does the rest.

Furthermore, $f(z)$ can have only one expansion (46) normally convergent in a sector contained in S. For the proof, we assume, multiplying z by a constant of modulus unity, if necessary, that the sector contains an infinite portion of the positive real axis. Then the normal convergence of (46) implies that the real part of σ_p approaches $-\infty$ as p increases. Hence, grouping finite sets of terms in (46), we can form a normally convergent series (25), which, of course, is unique. This shows that (46) is unique.

Let us see now what happens under the assumption that $f(z)$ is analytic throughout S, but not analytic throughout any sector containing S, with the same vertex as S and of greater opening than S.

There will be a subsector of S in which (46) converges absolutely. For (46) is obtained from the expansion (25) of $f(e^{i\phi_1}z)$ by replacing z by $e^{-i\phi_1}z$, and (46) is obtained similarly from $f(e^{i\phi_2}z)$.

On the other hand, in the expansion (25) of $f(e^{i\phi_1}z)$ (the μ's of the expansion will increase) there must be an $a(z)$ which contains an infinite number of terms. For a series (25) with all a's exponential polynomials, found by the Newton polygon procedure from (43), could be considered either as having increasing μ's or decreasing μ's. It would therefore converge normally in a sector above, and in a sector below, the positive real axis; thus, by (b) of §15, in a sector symmetric to the positive real axis. Hence $f(z)$ would be analytic in a sector containing S and of wider opening than S.

Thus some term in the expansion of $f(e^{i\phi_1}z)$ will give an infinite number of terms of (46). There will be an $h>0$ such that an infinite number of terms of (46) exceed $e^{-h|z|}$ in modulus as z increases with amp $z=\phi_1$. A similar state of affairs holds for amp $z=\phi_2$. Hence (46) is not normally convergent in any subsector of S.

20. Suppose that some branch $f(z)$ of a solution of (43) is analytic in a sector of opening greater than π. According to Hille, the domain in which (46) converges absolutely is convex, and (46) converges uniformly in any closed region interior to its domain of absolute convergence. (This is essentially (b) of §15 and the result of §11.)

Hence the expansion (46) of $f(z)$ will converge absolutely for every z, and uniformly in every finite area. Thus $f(z)$ is an integral function.

We shall prove that $f(z)$ is an exponential polynomial.

Since, in every sector of opening less than 2π, there is a subsector in which $f(z)$ has a normally convergent expansion (46), and since (46) is unique for any sector, $f(z)$ has a development (46) normally convergent all over the plane, that is, in a sector of opening 2π. Thus, the remainder after a large number of terms in (46) approaches zero uniformly all over the plane. As every remainder is an integral function, the remainder after a large number of terms is a constant, which has to be zero, for no $ve^{\sigma z}$ with $\sigma\neq 0$ and $v\neq 0$ can be a constant.

Thus $f(z)$ is an exponential polynomial. We shall prove that every exponent in $f(z)$ is a linear combination of the exponents of the P's in (43), with rational coefficients.

We have, identically in w and z,

$$P_n w^n + \cdots + P_0 = (w - f(z))(P_n w^{n-1} + Q_1 w^{n-2} + \cdots + Q_{n-1}), \tag{47}$$

with every Q an exponential polynomial.

From among the exponents present (effectively) in all of the P's, we select those whose real parts are a minimum, and from those just selected we pick that exponent which has a minimum coefficient for $(-1)^{1/2}$. Let β be the exponent thus selected. We write (47)

$$(48) \qquad e^{-\beta_n z}P_n w^n + \cdots + e^{-\beta_0 z}P_0 = (w - f(z))(e^{-\beta_n z}P_n w^{n-1} + \cdots + e^{-\beta_0 z}Q_{n-1}).$$

In the functions $e^{-\beta_i z}P_i$, those exponents which are not zero have either a positive real part or a zero real part and a positive coefficient for $(-1)^{1/2}$. Also zero is actually present as an exponent in certain of the functions.

If w is given a constant value, the first member of (48) becomes an exponential polynomial in z. We attribute to w such a constant value that the first member of (48) and also $w-f(z)$, when written as exponential polynomials, may each contain a constant term, that is, a term of exponent zero. This is clearly possible.

In $w-f(z)$, with w the constant taken above, we pick out those exponents whose real parts are a minimum, and from the exponents just selected we choose that one for which the coefficient of $(-1)^{1/2}$ is a minimum. Let γ be the exponent thus found. We write

$$w - f(z) = e^{\gamma z}(e^{-\gamma z}(w - f(z))).$$

By the first theorem of our paper* *A factorization theory for functions* $\sum a_i e^{\alpha_i z}$ every exponent in $e^{-\gamma z}(w-f(z))$ is linear in the exponents in the first member of (48) with rational coefficients. But since the exponent zero is present in $w-f(z)$, the exponent $-\gamma$ is present in $e^{-\gamma z}(w-f(z))$. Hence γ is linear in the exponents of the first member of (48), with rational coefficients. Then every exponent in $w-f(z)$ is linear in the exponents in the functions $e^{-\beta_i z}P_i$, with rational coefficients, and from what we know of β, this gives our result.

The results of the present article are given by the following two theorems:

THEOREM I. *If a solution w of* (1) *is a uniform function, more generally a function uniform in a sector of opening greater than π, then $w=Q/P_n$, where Q is an exponential polynomial whose exponents are linear combinations of the exponents in the P's, with rational coefficients.*

THEOREM II. *If P/Q, with P and Q exponential polynomials, is analytic in a sector of opening greater than π, then P/Q is an exponential polynomial.*†

21. We write (43)

$$(49) \qquad w^n + Q_{n-1}w^{n-1} + \cdots + Q_0 = 0,$$

where $Q_i=P_i/P_n$. We shall call (49) *irreducible* if its first member is not the product of two polynomials in w, both of degree less than n, with coefficients which are the quotients of two exponential polynomials.

* Ritt, (2).

† Cf. Ritt, (1). [Remark inserted in proof. The method of (1) can be extended so as to yield Theorem II.]

THEOREM III. *An irreducible equation* (49) *defines only a single analytic function.*

Consider any area in which the n solutions of (49) are analytic.

Suppose that $w_1, \cdots, w_r$, $r<n$, are distinct solutions of (49) which are continuations of each other, but of no other solution of (49). Then the symmetric functions of $w_1, \cdots, w_r$ are uniform all over the plane. Since the symmetric functions satisfy equations like (49), the symmetric functions are quotients of exponential polynomials. This shows that (49) is not irreducible.

We prove now

THEOREM IV. *Let* (49) *be irreducible, with* $n>1$. *Let* S *be any sector of opening greater than* π. *Then there exist in* S, *and outside an arbitrarily large circle, branch points of* w *which generate the entire group of monodromy of* w.

Suppose that those branch points which lie in S, and outside some circle, generate only a proper subgroup H of the group of monodromy, G. Construct a rational combination R of the solutions of (49) which is invariant (as a function of z) under H, but not under G. Then R is uniform in S. As R is a solution of an equation like (49), R is uniform all over the plane, hence invariant under G. This contradiction proves our theorem.

22. We consider an equation (43) for which the corresponding equation (49) is irreducible. A ray emanating from the origin will be called *singular* if there does not exist a sector, containing in its interior an infinite portion of the ray, in which all of the n branches of w are analytic.

We shall prove that w has at most a finite number of singular rays.

Let ϕ be any angle. Then, by section III, the n solutions of

$$P_n(e^{i\phi}z)w^n + \cdots + P_0(e^{i\phi}z) = 0$$

are analytic in two sectors, one below, one above the real axis, each with a side in the direction of the positive real axis. Then the ray amp $z=\phi$ cannot have singular rays making arbitrarily small angles with it. This proves that there are at most a finite number of singular rays.

According to Theorem II, if w is not an integral function, every sector with vertex at the origin, of opening greater than π, contains at least one singular ray. Hence, if w is not an integral function, it has at least two singular rays. When there are just two singular rays, the angle between them is π.

If w is not uniform, then no branch of w can be uniform in any sector of opening greater than π. In that case, w must have at least two singular rays such that, in any sector containing one of them, there are an infinite number of branch points of w. Singular rays of this type will be called *branch rays*.

Let w have $r\geqq 2$ singular rays, which divide the plane into sectors

$S_1, \cdots, S_r$. It is easy to see that in every S_i there is a subsector in which every branch of w is analytic. According to §19, every branch of w has a development (46), absolutely convergent in some subsector of S_i. If T_i is any sector with vertex at the origin every point of which, except the vertex, is interior to S_i, the development (46) of every branch is normally convergent in a subsector of T_i.

It will be shown later by examples that some of the branches of w may be analytic in a sector containing in its interior an infinite portion of a singular ray of w.

We summarize the principal results of §§19, 21 in the following theorem, in which it is understood that every Q is the quotient of two exponential polynomials.

THEOREM V. *Let w be the solution of the irreducible equation*

$$w^n + Q_{n-1}w^{n-1} + \cdots + Q_0 = 0. \tag{50}$$

Then w has a finite number of singular rays. Given any two adjacent singular rays, there is a subsector of the sector formed by them in which every branch of w is analytic. If w is not integral, it has at least two singular rays. If $n>1$, w has at least two branch rays, and no branch of w is uniform in a sector of opening greater than π. If a branch w_1 is analytic in a subsector of the sector S formed by two (not necessarily adjacent) singular rays, then w_1 can be expanded in a series

$$v_1 e^{\sigma_1 z} + \cdots + v_p e^{\sigma_p z} + \cdots ; \quad \lim |\sigma_p| = \infty ,$$

with constant v's, absolutely convergent in some subsector T of S. In any sector with the same vertex as T, and lying, except for its vertex, interior to T, the expansion is normally convergent. There is but one expansion of this type which converges normally to w_1 in a sector contained in S. If w_1 is not analytic in any sector with the same vertex as T, and containing T, except for the vertex, in its interior, the expansion is not normally convergent throughout T.

23. The singular rays of w admit of quite precise description in terms of the exponents in the P's of equation (1).

The poles of w are all located at the zeros of P_n. Those branch points of w which are not zeros of P_n are found by equating the discriminant of (1) to zero. The discriminant is a polynomial in the P's divided by a power of P_n. Hence the singular points of w are found by equating to zero two exponential polynomials, P_n and the numerator of the discriminant.

The distribution of the zeros of exponential polynomials has been investigated by Tamarkin, Wilder, and Pólya in papers to which reference

has already been made. Of their results, the following is the one of chief interest for us.

Consider any exponential polynomial P. Let the exponents in P be plotted in the complex plane, and let the smallest convex polygon $\mathfrak{A}$ which contains the exponents, in its interior or upon its boundary, be constructed. Let the sides of $\mathfrak{A}$ be $\sigma_1, \cdots, \sigma_p$. Let d_i be a ray which is the image, with respect to the real axis, of the normal to σ_i directed outwardly to $\mathfrak{A}$. Then there exist p half-strips, each parallel to, and directed like, some ray d_i, which contain all of the zeros of P. If the length of σ_i is s_i, the number of zeros in the strip parallel to d_i and less than r in modulus is equal to

$$\frac{\sigma_i r}{2\pi} + O(1).$$

The singular rays of w are parallel to the rays d_i, constructed for P_n and for the numerator of the discriminant. Of course, not every d_i need give a singular ray.

We shall show by means of an example that, along the distant part of a branch ray, the branches of w may break up into sets such that branches belonging to different sets are not connected by those branch points of w which cluster about the given branch ray. For instance, w may have a branch which is uniform in a sector containing an infinite portion of a branch ray of w.

Let w be defined by

$$w^3 + 3e^z w^2 + e^{(2+i)z} w + 1 = 0.$$

The exponents in the discriminant of w are found to be $6+2i$, $6+3i$, 0, 3 and $3+i$. The rays d_i for the discriminant are found to have inclinations 0, $\pi/2$, arctan $3/2$ (first quadrant), arctan 2 (third quadrant). When we seek the expansions (25), according to the method of III, for the three branches of w, we find one expansion with a's which are exponential polynomials, and two with a's which are infinite series. This means that one branch of w is uniform along the distant part of the positive real axis, while two branches are connected with each other by branch points of arbitrarily large modulus in a half-strip directed like the positive real axis.

We find a similar state of affairs along the ray of slope $3/2$, which is investigated by making the substitution $z=(2+3i)z'$. There is one uniform branch and the other two are connected.

Along the rays of slopes ∞ and 2, all of the branches are connected.

The following example shows that w may have an infinite number of poles along a singular ray and that still some branch of w may be analytic in a sector containing the distant part of the ray. Let w be defined by

$$(e^{iz} - 1)w^2 - 2e^z w - 1 = 0.$$

Then w has a pole for every integral value of z. Still one of the developments (25) of w has a's which are exponential polynomials. This is seen on writing

$$w = \frac{e^z \pm (e^{2z} + (e^{iz} - 1))^{1/2}}{e^{iz} - 1}$$

and expanding the radical in powers of $(e^{iz}-1)/e^{2z}$. Thus one of the branches of w is analytic in a sector containing an infinite portion of the positive real axis.

24. Let us imagine that some branch $f(z)$ of w is analytic in a half-plane, to fix our ideas, in an upper half-plane. Then the development (25) of $f(-z)$, with decreasing μ's, is found from that of $f(z)$ with increasing μ's by replacing z by $-z$. As both developments have decreasing λ's, both developments have only a finite number of λ's. Hence $f(z)$ is of the form

$$a_1(z)e^{\lambda_1 z} + \cdots + a_p(z)e^{\lambda_p z} \tag{51}$$

where each $a(z)$ is an algebraic combination of exponentials with pure imaginary exponents. It is evident that (51) will represent all of the branches of w, all over the plane. Hence we can remove a strip from the plane and leave two half-planes in which every branch of w is analytic.

Consider a w which has just three branch rays. From what precedes, we see that no two of them are extensions of each other. It follows from Theorem IV that the branch points which cluster about the distant portion of any branch ray connect up all of the branches of w.

References

Hille, *Note on Dirichlet series with complex exponents*, Annals of Mathematics, (2), vol. 25 (1924), pp. 261–278.

Pólya, *Geometrisches ueber die Verteilung der Nullstellen gewisser ganzer transzendenter Funktionen*, Münchener Berichte, 1920.

Ritt, (1) *On the zeros of exponential polynomials*, these Transactions, vol. 31 (1929), pp. 680–686.

(2) *A factorization theory for functions* $\sum_{i=1}^{n} a_i e^{\alpha_i z}$, these Transactions, vol. 29 (1927), pp. 584–596.

Tamarkin, *Some general problems in the theory of ordinary linear differential equations and expansion of an arbitrary function in series of fundamental functions*, Mathematische Zeitschrift, vol. 27 (1928), pp. 1–54. Also, earlier papers there mentioned.

Schwengler, *Geometrisches ueber die Verteilung der Nullstellen spezieller ganzer Funktionen*, Dissertation, Zurich, 1925.

Wilder, *Expansion problems in ordinary linear differential equations with auxiliary conditions at more than two points*, these Transactions, vol. 18 (1917), pp. 415–432.

Columbia University,
New York, N. Y.

ON THE ZEROS OF EXPONENTIAL POLYNOMIALS*

BY

J. F. RITT

By an exponential polynomial, we shall mean a function

$$a_0e^{\alpha_0 z} + \cdots + a_me^{\alpha_m z} \tag{1}$$

with constant a's and with constant α's distinct from one another. The distribution of the zeros of such functions, and of more general functions in which the a's are polynomials in z, rather than constants, has been investigated by Tamarkin, Pólya and Schwengler†. The very elegant results secured by them will be described, to some extent, below. The present writer has treated the question of factorizing an exponential polynomial into a product of exponential polynomials‡.

We present here two results. In §1, we prove that if every zero of one exponential polynomial is also a zero of a second exponential polynomial, the quotient of the second function by the first is an exponential polynomial. In §2, we study the function

$$1 + a_1e^{\alpha_1 z} + \cdots + a_me^{\alpha_m z},$$

with real α's such that

$$0 < \alpha_1 < \cdots < \alpha_m.$$

We consider any horizontal strip of the complex plane, and derive an expression for the sum of the real parts of those zeros of the exponential polynomial which are situated in the strip. The result obtained is analogous to the theorem that the product of the zeros of

$$1 + a_1z + a_2z^2 + \cdots + a_mz^m$$

is $(-1)^m/a_m$.

---

* Presented to the Society, October 27, 1928. Received by the editor of the Bulletin in November, 1928, accepted for publication in the Bulletin, and subsequently transferred to these Transactions.

† Tamarkin, Mathematische Zeitschrift, vol. 27 (1927), p. 1, and earlier papers there referred to; Pólya, Münchner Berichte, 1920; Schwengler, *Geometrisches ueber die Verteilung der Nullstellen* etc., Dissertation, Zurich, 1925.

‡ These Transactions, vol. 29 (1927), p. 584.

1. **Division.** THEOREM. *Let*

$$A(z) = a_0e^{\alpha_0 z} + \cdots + a_me^{\alpha_m z}, \qquad B(z) = b_0e^{\beta_0 z} + \cdots + b_ne^{\beta_n z}. \tag{2}$$

Suppose that $B(z) \not\equiv 0$, *and that* $A(z)/B(z)$ *is an integral function. Then there exists a*

$$C(z) = c_0e^{\gamma_0 z} + \cdots + c_pe^{\gamma_p z}$$

such that $A(z) = B(z)\ C(z)$.

We begin our proof by describing a result of Tamarkin, Pólya and Schwengler. Let the exponents α in $A(z)$ be plotted in the complex plane, and let the smallest convex polygon $\mathfrak{A}$ which contains them be constructed. Let the sides of $\mathfrak{A}$ be designated by

$$\sigma_1, \cdots, \sigma_l.$$

Let d_i, $i = 1, \cdots, l$, represent a ray which is the image, with respect to the real axis, of a perpendicular to σ_i erected exterior to $\mathfrak{A}$. It is proved by the above-named writers that there exist l half-strips,* each parallel to, and extending in the same direction as, one of the rays d_i, which contain all of the zeros of A. If s_i is the length of σ_i, the number of zeros in the half-strip parallel to d_i whose moduli are less than r is asymptotically equivalent to $rs_i/(2\pi)$.

Consider now the convex polygon $\mathfrak{B}$ corresponding to $B(z)$. As every zero of B is also a zero of A, it is clear, from the asymptotic formula for the number of zeros in a half-strip, that to every side τ of $\mathfrak{B}$ there corresponds a side of $\mathfrak{A}$ which is parallel to τ, at least as long as τ, and whose outward perpendicular has the same direction as the outward perpendicular to τ.

We shall suppose that the α's in (1) are so ordered that α_i comes before α_j if the real part of α_i is less than that of α_j, or if the real parts are equal but the coefficient of $(-1)^{1/2}$ in α_i is less than that in α_j. When the α's are real and non-negative, and $a_m \not\equiv 0$, we shall call α_m the degree of the function (1).

LEMMA. *If* $A(z)$, *and* $B(z) \not\equiv 0$, *are two exponential polynomials with real non-negative exponents, we have*

$$A = QB + R, \tag{3}$$

where Q *and* R *are two exponential polynomials with real non-negative exponents, and where* R, *if not zero, is of lower degree than* B.

If, in (2), we have $\alpha_m < \beta_n$, we have (3) with $Q = 0$, $R = A$. Suppose then that $\alpha_m \geqq \beta_n$.

* By a half-strip is meant an infinite region comprised between two half-lines and a line perpendicular to both of them.

Let $\alpha_m, \cdots, \alpha_{m-i}$ be those α's which are at least as great as β_n. Consider the exponential polynomial

$$C = A - \frac{(a_m e^{\alpha_m z} + \cdots + a_{m-i}e^{\alpha_{m-i}z})}{b_n e^{\beta_n z}} B, \tag{4}$$

whose exponents are non-negative.

Suppose first that B consists of one term. Then C is either zero, or of smaller degree than B, so that we have, in (4), a representation (3) with Q the fraction in the second member of (4), and $R=C$.

Suppose now that B has at least two terms. If C is not zero, its degree is either less than β_n or equal to $\alpha_m-(\beta_n-\beta_{n-1})$. If C is zero, or of degree less than β_n, we have, in (4), a representation (3). Otherwise, we repeat the process just described, subjecting C to the treatment received by A. As $\beta_n-\beta_{n-1}$ is a fixed quantity ,we arrive in a finite number of steps at a representation (3).

It is a simple consequence of the asymptotic formula for the distribution of the zeros of exponential polynomials that the representation (3) is unique.

We return now to the A and B of our theorem, whose exponents may, of course, be complex. We assume that $A \not\equiv 0$. Grouping together those terms of A whose exponents have like real parts, we write

$$A = P_1 e^{u_1 z} + \cdots + P_j e^{u_j z}, \tag{5}$$

where the u's are real, increasing with their subscripts, and where the P's are of the type

$$g_1 e^{v_1 iz} + \cdots + g_p e^{v_p iz}, \tag{6}$$

with real v's which increase with their subscripts.

As it does not disturb the zeros of A, or affect the divisibility problem which we are studying, to multiply A by an exponential, we suppose that u_1, and the smallest v in P_j are both zero.

Similarly, we suppose that

$$B = Q_1 e^{w_1 z} + \cdots + Q_h e^{w_h z}, \tag{7}$$

with stipulations identical with those made above for A.

The quantity u_j is the difference between the abscissas of any rightmost point and any leftmost point of $\mathfrak{A}$. Then because to every side of $\mathfrak{B}$ there corresponds a side of $\mathfrak{A}$ at least as long, and having the same direction, it is clear that $u_j \geqq w_h$.

Let us suppose, for the present, that B consists of at least two terms, that is, in (7), $h \geqq 2$, $Q_1, Q_h \not\equiv 0$.

Let $u_j, \cdots, u_{j-r}$ be those u's which exceed $u_j-(w_h-w_{h-1})$. We shall prove that the quotients of $P_j, \cdots, P_{j-r}$ by Q_h are exponential polynomials of the type (6).

If Q_h is a constant, this is certainly so. Suppose that Q_h is not a constant. Then $\mathfrak{B}$ has a right-hand vertical side whose length is the greatest v in Q_h. Then $\mathfrak{A}$ must have a right-hand side of at least the same length. That is, the greatest v in P_j is not less than that of Q_h. By the lemma above,

$$P_j = SQ_h + R$$

where S and R are of type (6), with non-negative v's, and where, if $R \not\equiv 0$, the greatest v in R is less than that of Q_h.

We say that R is zero. Suppose that this is not so. Then

$$A - SBe^{(u_j-w_h)z} = \cdots + Re^{u_jz}. \tag{8}$$

The terms which precede the last in the second member of (8) are products of polynomials (6) by exponentials e^{dz} with every $d \geqq 0$ and less than u_j. Now the first member of (8) has every zero of B. But the right hand vertical side of the polygon for the first member of (8) is shorter than the corresponding side of $\mathfrak{B}$. This shows that $R \equiv 0$.

If $u_{j-1} > u_j-(w_h-w_{h-1})$, we have

$$A - SBe^{(u_j-w_h)z} = \cdots + P_{j-1}e^{u_{j-1}z},$$

and it follows as above the P_{j-1} is the product of Q_h by a polynomial of type (6)*. Similarly, $P_{j-2}, \cdots, P_{j-r}$ are such products.

Consider now

$$D = A - \frac{P_je^{u_jz} + \cdots + P_{j-r}e^{u_{j-r}z}}{Q_he^{w_hz}}B.$$

If it is not identically zero, it is of the form

$$D = S_1e^{t_1z} + \cdots + S_ke^{t_kz},$$

where the S's are of type (6), where the t's are non-negative and increasing, and where $t_k \leqq u_j-(w_h-w_{h-1})$. If D is not zero, we can, since D has every zero of B, repeat the above procedure. As w_h-w_{h-1} is a fixed positive quantity, we can repeat our process only a finite number of times, so that, at some stage, we must reach a function like D above, which is zero. When that happens, we have A expressed as the product of B by an exponential polynomial.

When $h=1$, in (7), we have $B=Q_h$. It follows, as above, that every P

* The presence of negative v's in P_{j-1} would be of no significance.

is the product of B by a function of type (7). Our theorem is thus completely proved.

2. **Real exponents.** We deal with functions of the type

$$f(z) = 1 + a_1 e^{\alpha_1 z} + \cdots + a_m e^{\alpha_m z},$$

where the a's are any constants with $a_m \neq 0$, and the α's are any real numbers such that

$$0 < \alpha_1 < \cdots < \alpha_m.$$

Because $f(z)$ is close to unity when x ($z = x + yi$) is large and negative, and and close to ∞ when x is large and positive, there exist two vertical lines between which all of the zeros of $f(z)$ are comprised.

Let $R(u, v)$ be the sum of the real parts of those zeros of $f(z)$ for which $u < y < v$, where u and v are any real numbers with $v > u$. We shall prove that

$$R(u,v) = -\frac{(v-u)\log|a_m|}{2\pi} + O(1). \tag{9}$$

Let A be such that

$$|f(z) - 1| < 1 \tag{10}$$

for $x \leqq A$, and let $B > A$ be such that

$$\left|\frac{f(z)}{a_m e^{\alpha_m z}} - 1\right| < 1 \tag{11}$$

for $x \geqq B$. For any zero of $f(z)$, we have $A < x < B$.

Let S represent the sum of those zeros of $f(z)$ for which $u < y < v$. We assume that no zero of $f(z)$ lies on the lines $y = u$ or $y = v$. This assumption does not affect our results.

We have

$$2\pi i S = \int z \frac{f'(z)}{f(z)} dz, \tag{12}$$

the integration being performed in the positive sense around the rectangle of sides $x = A$, $x = B$, $y = u$, $y = v$.

Now

$$\int z \frac{f'(z)}{f(z)} dz = z \log f(z) - \int \log f(z) dz. \tag{13}$$

As $R(u, v)$ is the real part of S, we have to determine the imaginary part of the second member of (13).

We shall use that determination of $\log f(z)$ for which the coefficient of i at the point (A, v) is greater than $-\pi$ but not greater than π.

Let us determine first the variation of $z \log f(z)$ as z makes a circuit of the rectangle, starting from and returning to the point (A, v). Evidently $z \log f(z)$ is increased by $(A+vi)Ci$, where C is the variation in the amplitude of $f(z)$.

Because of (10), the variation of amp $f(z)$ along the side $x=A$ is less than π. To get an idea of the variation along $y=u$ and $y=v$, we consider that*

$$\text{amp } f(z) = \arctan \frac{Y}{X},$$

where X and Y are respectively the real and imaginary parts of $f(z)$. As z travels along a segment of the line $y=u$, for instance, amp $f(z)$ cannot undergo a variation as great as π unless X is zero at some point on the segment. Hence the variation of amp $f(z)$ along either horizontal side of our rectangle cannot exceed $\pi(p+1)$ where p is the number of zeros of X on such a side. On the line $y=u$, for instance,

$$X = 1 + b_1 e^{\alpha_1 x} + \cdots + b_m e^{\alpha_m x},$$

where the b's are real numbers depending upon u. It is known that a function like X cannot have more than n real zeros.† Hence the total variation of amp $f(z)$ along $y=u$, $y=v$, $x=A$, is less than $(2n+3)\ \pi$.

In virtue of (11), the variation of amp $f(z)$ along $x=B$ differs from the variation of the amplitude of $a_m e^{\alpha_m z}$ by less than π. The variation of the amplitude of $a_m e^{\alpha_m z}$ along $x=B$ is $\alpha_m(v-u)$. Hence the variation of amp $f(z)$ as z goes around the rectangle differs from $\alpha_m(v-u)$ by less than $(2n+4)\pi$.

The change in $z \log z$ is thus of the form

$$(A + vi)[\alpha_m(v - u) + O(1)]i.$$

The coefficient of i in this variation is, since the $O(1)$ is real,

$$A[\alpha_m(v - u) + O(1)]. \tag{14}$$

We shall estimate the imaginary part of the integral of $\log f(z)$. We put upon A the further condition that $\log f(z)$ have, for $x \leqq A$, an expression as an absolutely convergent Dirichlet series

$$\log f(z) = c_1 e^{\rho_1 z} + c_2 e^{\rho_2 z} + \cdots,$$

where the ρ's are positive and increase indefinitely. We see immediately that

* Tamarkin, loc. cit., p. 27–28, or Wilder, *Expansion problems* etc., these Transactions, vol. 18 (1917), pp. 415–447; pp. 420–427.

† Pólya and Szegö, *Aufgaben und Lehrsätze aus der Analysis*, vol. 2, p. 49.

$$\int_{v+Ai}^{u+Ai} \log f(z)dz = O(1). \tag{15}$$

We now take the side $y=u$. The amplitude of $f(z)$ at (A,u), differing by less than π from the amplitude at (A,v), does not exceed 2π in absolute value. As the variation of the amplitude along $y=u$ is less than $(n+1)\pi$ the absolute value of amp $f(z)$ is less than $(n+3)\pi$ on $y=u$. Hence the imaginary part of the integral along $y=u$ is less in absolute value than

$$(n+3)\pi(B-A). \tag{16}$$

We need the integral along $x=B$ of the *real* part of $\log f(z)$. We put upon B the restriction that, for $x \geqq B$, $\log f(z)$ admit an absolutely convergent development

$$\log f(z) = \alpha_m z + \log a_m + d_1 e^{\sigma_1 z} + d_2 e^{\sigma_2 z} + \cdots,$$

where the σ's are negative and decrease indefinitely. Thus the coefficient of i in the integral along $x=B$ is

$$\alpha_m B(v-u) + (v-u)\log|a_m| + O(1). \tag{17}$$

Finally, we must have the integral along $y=v$ of the imaginary part of $\log f(z)$. Along $y=v$, we have

$$|\operatorname{amp} f(z) - \alpha_m(v-u)| < (2n+4)\pi.$$

Hence

$$\left| \int_{v+Bi}^{v+Ai} \operatorname{amp} f(z)dz - \alpha_m(v-u)(A-B) \right| < (2n+4)\pi(B-A). \tag{18}$$

Understanding now that A and B are fixed, we have $A=O(1)$, $B-A=O(1)$, and find, from (14), (15), (16), (17) and (18), the expression (9) for $R(u,v)$.

COLUMBIA UNIVERSITY,
NEW YORK, N. Y.

ON THE SINGULARITIES OF LINEAR PARTIAL DIFFERENTIAL EQUATIONS*

BY

BERNARD OSGOOD KOOPMAN

The equations considered in this paper are linear systems of as many partial differential equations as there are unknown functions, of arbitrary order, and of coefficients analytic in the independent variables $(x_1, \cdots, x_n)$. We shall prove certain theorems regarding the region of existence of analytic solutions and the relation between the mobile singularities and the characteristics of the system. Our theorems have been proved in the case of one equation in two independent variables by Le Roux (Annales Scientifiques de l'École Normale Supérieure, (3), vol. 12 (1895), p. 227; Journal de Mathématiques, (5), vol. 4 (1898), p. 402), and in the real region under the same restrictions by Delassus (Annales Scientifiques de l'École Normale Supérieure, loc. cit., Supplement p. 53). They have been extended to equations in n independent variables by these authors, and by Hadamard (*Lectures on Cauchy's Problem*, Yale University Press, 1923, p. 72), but only under the severest restrictions, and by methods which do not admit of generalization.

We shall consider the system

$$L_i(u_1, \cdots, u_m) = 0 \qquad (i = 1, \cdots, m), \tag{1}$$

linear in the dependent variables u and their partial derivatives with respect to $x_1, \cdots, x_n$ up to the order r. The fact that r is the common highest order of the derivatives of all the u's in all the equations involves no restriction, for it may be brought about by means of suitable linear substitutions of the u's and the L's. The coefficients in (1) shall be analytic in $(x) = (x_1, \cdots, x_n)$ at the point $(x) = (x^0)$.

Let us pass briefly in review the classical treatment of the Cauchy problem for the system (1).

The object is to find m functions

$$u_i = f_i(x_1, \cdots, x_n) \qquad (i = 1, \cdots, m), \tag{2}$$

analytic at the point $(x) = (x^0)$, which, on substitution into (1), reduce it to an identity, and such that, on the preassigned analytic manifold through (x^0):

$$S(x_1, \cdots, x_n) = 0, \; S(x^0) = 0, \; \sum_{j=1}^{n} \left| \frac{\partial S}{\partial x_j} \right|_{(x^0)} \neq 0, \tag{3}$$

* Presented to the Society, October 27, 1928; received by the editors in February, 1929.

the functions (2) and their first $r-1$ derivatives take on the preassigned analytic values

$$\frac{\partial^{\lambda} u_i}{\partial x_1^{\lambda_1} \cdots \partial x_n^{\lambda_n}} = \overset{i}{\phi}_{\lambda_1 \cdots \lambda_n}(x_1, \cdots, x_n), \tag{4}$$
$$(0 \leqq \lambda = \lambda_1 + \cdots + \lambda_n \leqq r-1\ ;\ \lambda_j = 0,1, \cdots, r-1).$$

Here the 0th derivative means the function undifferentiated. The ϕ's must, of course, satisfy certain conditions of compatibility; thus, for example, we must have for all (x) on (3):

$$d\overset{i}{\phi}_{0 \cdots 0} = \overset{i}{\phi}_{1 \cdots 0} dx_1 + \cdots + \overset{i}{\phi}_{0 \cdots 1} dx_n. \tag{5}$$

By hypothesis one of the derivatives of S—let us say $\partial S/\partial x_n$—is not zero at (x^0). The transformation

$$\begin{aligned} x_1' &= x_1, \cdots, x_{n-1}' = x_{n-1}, \\ x_n' &= S(x_1, \cdots, x_n) \end{aligned} \tag{6}$$

is, then, analytic and of analytic inverse near (x^0). It transforms (1) and (4) into equations of the same form, but with (3) replaced by

$$x_n' = 0. \tag{7}$$

The new conditions of compatibility, (5) etc., will not contain x_n or dx_n', and indeed (4) and (5) etc., may be replaced by the following simpler initial data, in which the $\phi_{\lambda}'^{\,i}$'s are freed from all relations of the type (5):

$$\left(\frac{\partial^{\lambda} u_i}{\partial x_n'^{\lambda}}\right)_{x_n'=0} = \phi_{\lambda}'^{\,i}(x_1', \cdots, x_{n-1}')\ ; \tag{8}$$
$$(\lambda = 0,1, \cdots, r-1\ ;\ i = 1, \cdots, m).$$

The new equations obtained from (1) by the substitution (6) constitute m linear equations in the m quantities $\partial^r u_i/\partial x_n'^{\,r}$ for which they may be solved, at least formally, so that they yield

$$\frac{\partial^r u_i}{\partial x_n'^{\,r}} = l_i \qquad (i = 1, \cdots, m). \tag{9}$$

The l's are linear polynomials in the u's and their derivatives other than those appearing on the left; each of their coefficients is the quotient of a polynomial in the coefficients of the L's and in S and its derivatives, divided by a certain determinant D; hence the equations (9) exist effectively, and their coefficients are analytic at $(x'^0) \sim (x^0)$, provided $D \neq 0$ at this point.

The determinant D is found on computation to be a homogeneous poly-

nomial of degree mr in the first derivatives of S, the coefficients being themselves homogeneous of degree mr in the coefficients of the L's. The determinant D, regarded as a differential operator, shall be denoted by $D(S)$. In general, if the analytic manifold $F(x_1, \cdots, x_n) = 0$ is such that

$$\sum_i \left| \partial F/\partial x_j \right| \not\equiv 0 \quad \text{and} \quad D(F)_{F=0} = 0,$$

this manifold is said to be a *characteristic* of the system (1). If the two preceding relations hold merely at (x^0), the manifold is said to be tangent at this point to a characteristic.

We shall complete the solution under the supposit on that (3) is not tangent at (x^0) to a characteristic, i.e., $D(F)_{(x^0)} \neq 0$, and that the equations of the problem are reduced to the form (9) and (8); and, further, that $(x'^0) = (0)$,—i.e., x_j' is replaced by $x_j' - x_j^0$.

By introducing certain derivatives of u_i as further unknowns, our equations are replaced by an increased number of equations in an equal number of unknown functions, but involving only derivatives of the first order. Thus, after a change of notation, (9) and (8) are replaced by the equations

$$\frac{\partial u_i}{\partial x_m} = \sum_{\alpha=1}^{m} \sum_{\beta=1}^{n-1} A_{\alpha\beta}^{\ i} \frac{\partial u_\alpha}{\partial x_\beta} + \sum_{\alpha=1}^{m} B_\alpha^{\ i} u_\alpha + C^i \quad (i = 1, \cdots, m); \tag{10}$$

$$(u_i)_{x_n=0} = \phi_i(x_1, \cdots, x_{n-1}). \tag{11}$$

For the details of this reduction, cf., e.g., Goursat and Bourlet, *Leçons sur l'Intégration des Équations aux Dérivées Partielles du Premier Ordre*, Paris, Hermann, 1891, Chap. I.

We shall assume that in (10) $B \equiv C \equiv 0$. This entails no loss of generality; for consider the Cauchy problem with the differential equations

$$\frac{\partial U_i}{\partial x_n} = \sum_{\alpha=1}^{m} \sum_{\beta=1}^{n-1} A_{\alpha\beta}^{\ i} \frac{\partial U_\alpha}{\partial x_\beta} + \sum_{\alpha=1}^{m} x_0 B_\alpha^{\ i} \frac{\partial U_\alpha}{\partial x_0} + x_0 C^i \frac{\partial U_0}{\partial x_0},$$

$$\frac{\partial U_0}{\partial x_n} = 0 \qquad (i = 1, \cdots, m),$$

and the following conditions for $x_n = 0$:

$$U_i = x_0 \phi_i(x_1, \cdots, x_{n-1}), \quad U_0 = x_0,$$

x_0 being a new independent variable. This system is of the homogeneous form; and it is equivalent to (10) (11). For if u_i is a solution of the latter, $U_i = x_0 u_i$, $U_0 = x_0$, is a solution of the former, and conversely. The first statement is proved by substitution; the second results from the computation,

by means of the above equations, of the coefficients in the formal expansion of U_i in powers of the x's.

We have, then, to consider the system

$$\frac{\partial u_i}{\partial x_n} = \sum_{\alpha=1}^{m}\sum_{\beta=1}^{n-1} A^{i}_{\alpha\beta}\frac{\partial u_\alpha}{\partial x_\beta} \qquad (i = 1, \cdots, m), \tag{12}$$

along with the Cauchy data (11). The problem is solved by the *calcul des limites* of Cauchy. First, it is shown that (11) and (12) suffice exactly to determine the formal expansions of the functions u_i in power series in $x_1, \cdots, x_n$. Then it is proved that these formal series converge in the domain of convergence of solutions of a second system like (11) (12), in which ϕ_i and $A^{i}_{\alpha\beta}$ are replaced by power series whose coefficients are positive and greater than the absolute values of the corresponding coefficients in ϕ_i and $A^{i}_{\alpha\beta}$ of the given equations (11), (12) (cf. Goursat-Bourlet, loc. cit.).

It is at this point that we depart from the classical treatment, inasmuch as our special assumption of the linearity of the system (1) enables us to make use of dominating equations whose coefficients do not contain the unknown functions themselves (contrast: Goursat-Bourlet, loc. cit., p. 6: (2); our A's contain only (x), while in this reference they are functions of (u)).

By hypothesis, the functions $A^{i}_{\alpha\beta}$ and ϕ_i are analytic in and on the boundary of a certain neighborhood of $(x) = (0)$:

$$|x_j| \leqq a \qquad (j = 1, \cdots, n). \tag{13}$$

Hence, for all (x) in (13), and for a suitable $M > 0$ and $N > 0$,

$$|A^{i}_{\alpha\beta}| < M, \quad |\phi_i| < N.$$

We shall, accordingly, make use of the following dominating equations:

$$\frac{\partial v_i}{\partial x_n} = \frac{M\sum_{\alpha=1}^{m}\sum_{\beta=1}^{n-1}\frac{\partial v_\alpha}{\partial x_\beta}}{\left(1 - \frac{x_1 + \cdots + x_{n-1}}{a}\right)\left(1 - \frac{x_n}{a}\right)}, \tag{14}$$

$$(v_i)_{x_n=0} = \frac{N}{1 - \frac{x_1 + \cdots + x_{n-1}}{a}}.$$

To solve these, we set $x_1 = \cdots = x_{n-1} = y$, $x_n = z$, $v_1 = \cdots = v_m = v$; the equations become

$$\frac{\partial v}{\partial z} - \frac{Mm(n-1)a^2}{[a-(n-1)y][a-z]}\frac{\partial v}{\partial y} = 0, \quad (v)_{z=0} = \frac{Na}{a-(n-1)y}. \tag{15}$$

The first is reduced in the familiar manner to an ordinary differential equation, which is integrated by the separation of variables. The solution of the equations is found to be

$$v = Na\left\{[a-(n-1)y]^2 + 2Mm(n-1)^2a^2\log\left(1-\frac{z}{a}\right)\right\}^{-1/2}; \tag{16}$$

—more precisely, that branch of the radical which reduces to $Na/[a-(n-1)y]$ when $z=0$. The formal solution of (12) is analytic in a neighborhood $|y|<b$, $|z|<b(b\leqq a)$, determined by the conditions that (16) shall be analytic within it. But the radicand in (16) depends only on M and a and not on N. Hence (14) admits a solution analytic in $|x_j|<b$, b independent of N.

These results insure the existence of a solution of the Cauchy problem (11) (12), analytic in the neighborhood $|x_j|<b$, where b depends on the initial data ϕ_i only insofar as ϕ_i must be known to converge for $|x_j|<a$. Furthermore, the substitutions by means of which the general Cauchy problem (1) (3) (4) was reduced to the form (11) (12) involve in no wise the functions $\phi^i_{\lambda_1\cdots\lambda_n}$. From whence follows

THEOREM I. *If the coefficients of L_i and the function S are analytic in the neighborhood*

$$|x_j - x_j^0| < \alpha \qquad (j=1,\cdots,n), \tag{17}$$

throughout which $D(S)>$ const. >0, there will exist a certain fixed neighborhood

$$|x_j - x_j^0| < \beta \qquad (j=1,\cdots,n) \tag{18}$$

in which the Cauchy problem for $L_i=0$ and $S=0$ has an analytic solution. In the problem in question, the Cauchy data on $S=0$ are arbitrary, subject merely to the restriction of being analytic at all points of (17) *lying on $S=0$.*

THEOREM II. *If the coefficients of L_i and the function S are analytic at the point $(x)=(X)$, and if $D(S)_{(X)}\neq 0$, there will exist a closed neighborhood*

$$|x_j - X_j| \leqq \gamma \qquad (j=1,\cdots,n), \tag{19}$$

and a constant $\beta>0$, such that, if (x^0) be taken arbitrarily in (19), *and if analytic Cauchy data be assigned on that portion of the manifold $S(x)=S(x^0)$ which lies in the larger neighorhood*

$$|x_j - X| < 2\gamma \qquad (j=1,\cdots,n), \tag{20}$$

the corresponding Cauchy problem will admit a solution analytic in (18). *It is to be emphasized that β is independent both of (x^0) and of the Cauchy data.*

For the proof, we have but to take γ so small that throughout (20) L_i

and S are analytic, and $D(S) > \text{const.} > 0$. Then if (x^0) is taken such that $|x_j^0 - X| \leqq \gamma$, and $\alpha = \gamma$, Theorem I will apply. Since this furnishes a positive β for each point (x^0) of the closed neighborhood (19), we have but to take for the β of the present theorem their lower limit, which is positive.

When the process of analytic continuation is applied in an obvious manner, the following theorem is obtained:

THEOREM III. *Let R_{2n} be a $2n$-dimensional region of the complex $x_1 \cdots x_n$-space; let L_i have analytic coefficients in R_{2n}; finally, let S_{2n-2} be a $(2n-2)$-dimensional region of the analytic manifold $S(x_1, \cdots, x_n) = 0$, where S_{2n-2} lies in R_{2n}, and at no point of S_{2n-2} does $S(x_1, \cdots, x_n)$ satisfy the equation of characteristics $D(S) = 0$. There will then exist a $2n$-dimensional subregion P_{2n} of R_{2n}, containing S_{2n-2}, and such that, whatever be the analytic data on S_{2n-2}, the Cauchy problem will admit a solution analytic throughout P_{2n}.*

In this theorem the coefficients of L_i as well as the Cauchy data may be multiple-valued.

THEOREM IV. *Let the coefficients of L_i be analytic in the $2n$-dimensional region R_{2n}. If $u_1, \cdots, u_n$ are solutions of $L_i = 0$ which are analytic in R_{2n} except at all points of an analytic manifold S_{2n-2} lying in R_{2n}, then S_{2n-2} is a characteristic manifold of $L_i = 0$.*

Suppose the theorem false. Then we could find a point (X) on S_{2n-2} in a sufficiently small neighborhood of which S_{2n-2} is represented by the equation $S(x) = S(X)$, where $S(x)$ is analytic, and $D(S) \neq 0$, at (X). Let Theorem II be applied to L_i, S, (X), and let the point (x^0) be so chosen that

$$|x_j^0 - X| < \frac{\beta}{2} \qquad (j = 1, \cdots, n), \tag{21}$$

and that $S(x^0) \neq S(X)$. Evidently the functions $u_1, \cdots, u_n$ are analytic at those points of the manifold $S(x) = S(x^0)$ which lie in (20), for it is an easy matter to insure that all our points be in R_{2n}. Hence, by Theorem II, analytic solutions of $L_i = 0$ exist in (18) which take on the values of $u_1, \cdots, u_n$ and of appropriately chosen ones of their derivatives on the manifold $S(x) = S(x^0)$. By the uniqueness of the analytic solution of Cauchy's problem, these solutions are identical with the original solutions $u_1, \cdots, u_n$. But, by (21), the neighborhood of their analyticity (18) contains (X). But this is contradictory to the hypothesis that (X), which is on S_{2n-2}, is a singularity of $u_1, \cdots, u_m$. The truth of the theorem follows.

It is to be emphasized that these results are not in general true when stated for the real domain. Thus, if Theorem I were true for Laplace's

equation in real space, there could be no such thing as a solution with an isolated point singularity. As for Theorem IV in the real domain, even with equations with real characteristics, we may consider the equation

$$\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}-\frac{\partial^2 u}{\partial z^2}=0,$$

and the solution $u=f(x, y)$, where $f(x, y)$ satisfies

$$\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}=0$$

and admits the circle $x^2+y^2=1$ as a locus of essential singularities, without, however, being singular near this circle. This gives an example of a function satisfying the hypotheses of Theorem IV for reals, with S_{2n-2} as the cylinder $x^2+y^2=1$, which singularity is not a characteristic: if $S=x^2+y^2-1$,

$$\left(\frac{\partial S}{\partial x}\right)^2+\left(\frac{\partial S}{\partial y}\right)^2-\left(\frac{\partial S}{\partial z}\right)^2=4S+4.$$

COLUMBIA UNIVERSITY,
NEW YORK, N. Y.

A GENERALIZATION OF DIRICHLET'S SERIES AND OF LAPLACE'S INTEGRALS BY MEANS OF A STIELTJES INTEGRAL*

BY
D. V. WIDDER

INTRODUCTION

When one seeks to generalize a Taylor's series

$$\sum_{n=0}^{\infty} a_n z^n, \tag{1}$$

a natural method of procedure is to replace the set of integers n that appear as exponents by a more general set of numbers,

$$0 \leqq \lambda_0 < \lambda_1 < \lambda_2 < \cdots, \quad \lim_{n=\infty} \lambda_n = \infty . \tag{2}$$

If, however, λ_n is not an integer, z^{λ_n} is a multiple-valued function, and complications arise. This difficulty is easily met by making the transformation $z = e^{-s}$, which transforms the circle of convergence of (1) into a half plane. In this way one is led to Dirichlet's series

$$f(s) = \sum_{n=0}^{\infty} a_n e^{-\lambda_n s}. \tag{3}$$

It is only natural to proceed further and to replace the discrete set (2) by a continuous set, replacing λ_n by a variable t, which may vary from zero to infinity, and the sign of summation by the integral sign. The result is a function of the form

$$f(s) = \int_0^{\infty} a(t) e^{-st} dt. \tag{4}$$

Functions of this type were first studied by Laplace† and Abel,‡ who designated the function $f(s)$ as the generating function of $a(t)$, and $a(t)$ as the determining function of $f(s)$.

* Presented to the Society in part February 25, 1928, under the title *The singularities of functions defined by the Stieltjes integral* $\int_0^{\infty} e^{-st} d\alpha(t)$; in part March 29, 1929, under the title *On the composition of singularities of functions defined by factorial series*; received by the editors in May, 1929.

† Laplace, *Théorie Analytique des Probabilités*, Paris, 1812.

‡ Abel, Oeuvres, 2d edition, 1881, vol. 2, p. 67.

By the introduction of the Stieltjes integral

$$f(s) = \int_0^\infty e^{-st} d\alpha(t), \tag{5}$$

functions of types (3) and (4) may be considered simultaneously. Moreover, this integral serves to generalize both (3) and (4) since it includes a class of functions not included in either. It was shown by M. Fréchet* that an integral of type (5) with $\alpha(t)$ a function of bounded variation may be decomposed into the sum of three terms

$$f(s) = \int_0^\infty e^{-st} a(t) dt + \sum \alpha_n e^{-\lambda_n s} + \int_0^\infty e^{-st} du(t), \tag{6}$$

where $a(t)$ is a summable function, the λ_n are the points of discontinuity of $\alpha(t)$ with $\alpha_n = \alpha(\lambda_n + 0) - \alpha(\lambda_n - 0)$, and $u(t)$ is a continuous function of bounded variation which has a derivative zero except at a set of measure zero. In this way it is seen that (5) is more general than (3) or (4) from two points of view. The first integral in (6) corresponds to (4), but the summation, although it may be a Dirichlet's series, is not so in general. For, the points of discontinuity λ_n of $\alpha(t)$ may lie at a denumerable set of points which can not be arranged in the order (2), as for example at the rational points. Again, when the last term of (6) is present, $f(s)$ is on this account different in nature from either (4) or (5). For example, if $\alpha(t)$ is a continuous function for which there exists an everywhere dense set of non-overlapping intervals each one of which is a line of invariability for the function,† (6) defines a distinctly new type of function.

The integral (5) may in certain cases be transformed into a Riemann integral by the familiar formula for integration by parts.‡ If R is any positive number.

$$\int_0^R e^{-st} d\alpha(t) = e^{-sR}\alpha(R) - \alpha(0) + s\int_0^R e^{-st}\alpha(t) dt,$$

and

$$\int_0^\infty e^{-st} d\alpha(t) = \lim_{R=\infty} e^{-sR}\alpha(R) - \alpha(0) + s\int_0^\infty e^{-st}\alpha(t) dt$$

---

* M. Fréchet, *Sur les fonctionnelles linéaires et l'intégrale de Stieltjes*, Comptes Rendus du Congrès des Sociétés Savantes en 1913, pp. 45–54.

† E. W. Hobson, *The Theory of Functions of a Real Variable and the Theory of Fourier's Series*, 2d edition, 1921, vol. 1, p. 344.

‡ E. W. Hobson, loc. cit., vol. 1, p. 507.

provided that the integral on the right hand side exists or that the indicated limit exists. This equation shows that a study of the integral (5) and that of

$$s\int_0^\infty e^{-st}\alpha(t)dt \tag{7}$$

are not equivalent. Either integral may converge for a function $\alpha(t)$ which makes the other diverge, as the following examples show.

Define $\alpha(t)$ by the equations

$$\begin{aligned} &\alpha(t) = 0, && 0 \leqq t < 1, && \mu_n < t < n+1, \\ &\alpha(t) = e^{n^2}, && n \leqq t \leqq \mu_n. \end{aligned} \tag{8}$$

Here μ_n is defined by the relation

$$e^{n^2}[e^{-n} - e^{-\mu_n}] = 1/2^n \qquad (n = 1,2,3,\cdots).$$

Since

$$1 - e^{-n^2+n}2^{-n} \geqq 1/2 \qquad (n = 1,2,3,\cdots),$$

it is easily seen that

$$n < \mu_n < n+1,$$

so that equations (8) define $\alpha(t)$ without ambiguity. It is sufficient for our present purposes to suppose s real.

With this determination of $\alpha(t)$, (5) reduces to

$$\sum_{n=1}^{\infty}(-1)^{n-1}u_n, \quad u_{2m-1} = e^{m^2-ms}, \quad u_{2m} = e^{m^2-\mu_m s},$$

a series which clearly diverges for all s, since its general term does not approach zero. On the other hand (7) becomes

$$\sum_{n=1}^{\infty}e^{n^2}[e^{-ns} - e^{-\mu_n s}].$$

By the definition of μ_n this series is seen to reduce to the convergent series

$$\sum_{n=1}^{\infty}1/2^n$$

for $s=1$. It is not difficult to show that (7) also converges for all real s greater than unity. Thus (7) may converge when (6) diverges.

The opposite situation is illustrated by taking

$$\begin{aligned} &\alpha(t) = \sum_{m=1}^{n}e^{-2m}, \quad n \leqq t < n+1, \\ &\alpha(t) = 0, \quad 0 \leqq t < n. \end{aligned}$$

For this determining function (6) becomes

$$\sum_{n=1}^{\infty} e^{-2n} e^{-ns},$$

a series which evidently converges for $s > -2$. But (7) is divergent when s is negative. For, since $\alpha(t)$ is a monotonic increasing function we have

$$\alpha(t)e^{-st} \geqq \alpha(1)e^{-st} = e^{-2}e^{-st}, \qquad t \geqq 1.$$

The right-hand side of this inequality becomes infinite as t becomes infinite if s is negative, so that (7) can not converge. Thus (6) may converge when (7) diverges. Consequently we shall generally treat the Stieltjes integral directly without appeal to the corresponding Riemann integral. Moreover, the results obtained in this way are more compact, better suited to the applications to which they are put.

The chief purpose of the present paper is to discuss the effect of the determining function on the singularities of the generating function, and in particular to obtain a result for the composition of singularities analogous to the familiar theorem of Hadamard for Taylor's series. In order to obtain such results it is found necessary to study the fundamental properties of the functions (5). It is found that many of the familiar properties of Dirichlet's series are common to these functions, as is to be expected. For example, the region of convergence is a half plane or the whole plane; a half plane of absolute convergence may or may not exist. A discussion of the rate of increase or decrease of the generating function $f(s)$ as s recedes to infinity along lines parallel to the axis of imaginaries is necessary for subsequent developments. It is seen that $f(s)$ can not increase more rapidly than in the case of Dirichlet's series. An expression for the determining function in terms of the generating function is next obtained. Fractional derivatives and integrals of the determining function are also obtained by similar formulas. Part I closes with a proof that the product of two generating functions is itself a generating function in certain cases.

Part II begins with a proof that if $\alpha(t)$ is monotonic, then $f(s)$ has a singularity at the real point of the axis of convergence. If the Stieltjes integral reduces to a power series, this reduces to a familiar result concerning power series with positive coefficients. In the next section the most important result of the paper is obtained. In its simplest form it states that if the function $f(s)$ defined by (5) has singularities at the points α and if the function

$$\phi(s) = \int_0^{\infty} e^{-st} d\beta(t)$$

has singularities at the points β, then

$$F(s) = \int_0^\infty e^{-st}\alpha(t)d\beta(t)$$

has singularities at most at the points $\alpha+\beta$ and β under certain conditions imposed on the rate of increase of $f(s)$ and $\phi(s)$ on vertical lines and upon the distribution of the singularities α and β. The result reduces to Hadamard's if the functions $\alpha(t)$ and $\beta(t)$ are step functions with discontinuities at the integral points and to a result of the author for Dirichlet's series if the discontinuities are at a set of points (2). After developing certain sufficient conditions that a function $f(s)$ can be expressed as a generating function it is shown that Hurwitz's result regarding the addition of singularities of power series is also included in the above result. Generalizations of familiar theorems of Faber and Leau are also obtained, and generating functions for which the corresponding determining function has special form are treated. For example, the case in which the determining function is itself a generating function is of particular interest, since the function $\Gamma(s)$ is of this nature. An application of the result of the paper is made to functions defined by factorial series. A necessary and sufficient condition that a function $f(s)$ can be developed into such a series is known. The condition demands that the function $f(s)$ be a generating function of specified type. Hence it is possible to discuss the composition of singularities of such functions. It is found that if

$$f(s) = \sum_{n=0}^{\infty} \frac{a_n n!}{s(s+1)\cdots(s+n)}$$

has singularities at points α, and if

$$\phi(s) = \sum_{n=0}^{\infty} \frac{b_n n!}{s(s+1)\cdots(s+n)}$$

has singularities at points β, then

$$F(s) = \sum_{n=0}^{\infty} \frac{(a_0 b_n + a_1 b_{n-1} + \cdots + a_n b_0)n!}{s(s+1)\cdots(s+n)}$$

has singularities at most at the points $\alpha+\beta$ under suitable restrictions. The similarity of this result with that of Hurwitz is apparent.

It is thus seen that the introduction of the Stieltjes integral does much toward the unification of the theory of functions of a complex variable, since by it power series, Dirichlet's series, factorial series, the generating functions of Laplace, etc. may all be treated together. Functions defined as differently as $\Gamma(s)$ and $\zeta(s)$, for example, come to be special cases of a single theory.

PART I. THE FUNDAMENTAL PROPERTIES OF THE GENERATING FUNCTION

1. **The region of convergence.** Let $\alpha(t)$ be a complex function of the real variable t of bounded variation in every interval $0 \leqq t \leqq t_1$, t_1 being arbitrarily large. In order to simplify certain subsequent formulas we assume further that $\alpha(0)=0$. If $\alpha(0)$ were not zero, a redefinition to make it so would amount only to adding a constant to the function $f(s)$ defined by (5), so that no essential change in the properties of $f(s)$ is effected. Under these conditions it is a familiar fact that the Stieltjes integral

$$\int_0^{t_1} F(t)d\alpha(t)$$

exists for every continuous function $F(t)$.* Let s be a complex variable, $s=\sigma+i\tau$. Then the integral

$$S(t_1,s) = \int_0^{t_1} e^{-st}d\alpha(t) \tag{1.1}$$

exists for all values of s and for all values of $t_1>0$. $S(0, s)$ is defined to be zero. We wish first to discuss the conditions of convergence of the corresponding improper integral obtained by allowing t_1 to become infinite in (1.1),

$$f(s) = \int_0^{\infty} e^{-st}d\alpha(t). \tag{1.2}$$

Following Laplace we shall designate the function defined by this integral when it converges as the *generating* function and the function $\alpha(t)$ as the *determining* function. We prove first

THEOREM 1. *If the integral* (1.2) *converges for a value* $s_0=\sigma_0+i\tau_0$ *of* s, *then it converges for all values of* s *for which* $\sigma>\sigma_0$.

Since (1.2) converges for $s=s_0$, the function $S(t_1, s_0)$ defined by (1.1) approaches a limit as t_1 becomes infinite. Consequently there exists a constant A independent of t_1 in the interval $0 \leqq t_1 < \infty$ such that

$$|S(t_1,s_0)| < A, \qquad 0 \leqq t_1 < \infty .$$

Let $s_1=\sigma_1+i\tau_1$ be an arbitrary complex number for which $\sigma_1>\sigma_0$, and set $h=s_1-s_0$. Then the real part of h, $\sigma_1-\sigma_0$, is positive. We wish to show that

$$\lim_{t_1=\infty} \int_0^{t_1} e^{-(s_0+h)t}d\alpha(t)$$

* E. W. Hobson, loc. cit., vol. 1, p. 506 et seq.

exists. In order to prove this we note that*

$$\int_0^{t_1} e^{-s_0 t} e^{-ht} d\alpha(t) = \int_0^{t_1} e^{-ht} dS(t, s_0),$$

and apply the formula for integration by parts,

$$\int_0^{t_1} e^{-s_0 t} e^{-ht} d\alpha(t) = S(t_1, s_0) e^{-ht} + h \int_0^{t_1} S(t, s_0) e^{-ht} dt. \tag{1.3}$$

But

$$| S(t, s_0) e^{-ht} | < A e^{-(\sigma_1 - \sigma_0) t},$$

and the limit of the first term on the right-hand side of (1.3) is seen to be zero. The second term also approaches a limit since

$$\left| \int_0^{t_1} S(t, s_0) e^{-ht} dt \right| < \int_0^{t_1} A e^{-(\sigma_1 - \sigma_0) t} dt,$$

and since the integral

$$A \int_0^{\infty} e^{-(\sigma_1 - \sigma_0) t} dt$$

converges. It follows that (1.2) converges for $\sigma > \sigma_0$. It is important to observe that the transformation (1.3) has enabled us to replace the integral (1.2) which is not in general absolutely convergent† by an absolutely convergent integral:

$$\int_0^{\infty} e^{-st} d\alpha(t) = h \int_0^{\infty} S(t, s_0) e^{-ht} dt. \tag{1.31}$$

The transformation (1.3) reduces to the transformation of Abel when the Stieltjes integral becomes a series.

As an immediate consequence of Theorem 1 it follows that the divergence of (1.2) for a point $s_0 = \sigma_0 + i\tau_0$ implies its divergence at all points for which $\sigma < \sigma_0$. Consequently the same possibilities arise here as in the case of Dirichlet's series: (a) the integral may converge for all values of s; (b) it may converge for no value of s; (c) there may exist a constant σ_c such that the integral converges for $\sigma > \sigma_c$ and diverges for $\sigma < \sigma_c$. In case (c) the line $\sigma = \sigma_c$ is called the *axis of convergence* and the half plane $\sigma > \sigma_c$ the *half plane of convergence.*

* See, for example, T. Carleman, *Sur les Équations Intégrales Singulières à Noyau Réel et Symétrique*, p. 11, Theorem III.

† The integral (1.2) is said to be absolutely convergent if $\int_0^{\infty} |e^{-st}| du(t)$ converges, where $u(t)$ is the total variation of $\alpha(t)$ from zero to t. The definition of the total variation of a complex function of the real variable is exactly the same as that of a real function, no separation into real and imaginary parts being necessary.

In a similar way one defines the *axis of absolute convergence* and the *region of absolute convergence.*

Theorem 1 enables us to discuss the region of convergence of an integral of the form

$$\int_{-\infty}^{\infty} e^{-st}d\alpha(t) \tag{1.4}$$

where $\alpha(t)$ is now considered to be of finite variation in *every* finite interval. By the transformation $s=-s'$, the integral

$$\int_{-\infty}^{0} e^{-st}d\alpha(t) = \int_{0}^{\infty} e^{st}d\alpha(-t) \tag{1.5}$$

becomes one of type (1.2). Consequently the region of convergence of (1.5) is a half plane lying to the left of a line $\sigma=\text{const.}$ Hence the region of convergence of (1.4), which is the analogue of Laurent's series, is a strip of the plane or obvious modifications of such a strip. For example, in the case of the function

$$\Gamma(s) = \int_{0}^{\infty} t^{s-1}e^{-t}dt = \int_{-\infty}^{\infty} e^{-st}e^{-e^{-t}}dt$$

the strip of convergence becomes the half plane $\sigma>0$. On the other hand the region of convergence may reduce to a straight line or a set of points on a line, or it may disappear completely. That these cases may actually occur is evident in view of the fact that power series form special cases of the integral (1.2).

2. **Uniform convergence.** We prove the following theorem:

THEOREM 2. *If the integral* (1.2) *converges for* $s=s_0$, *and if* H *is any positive number, then it converges uniformly in the region*

$$|s-s_0| \leqq (\sigma-\sigma_0)He^{H(\sigma-\sigma_0)}, \quad \sigma \geqq \sigma_0.$$

Let ϵ be an arbitrarily small positive number. Then it is possible to determine a number T greater than H and so large that

$$|S_{t_1}(t,s_0)| = \left|\int_{t_1}^{t} e^{-s_0t}d\alpha(t)\right| < \epsilon, \quad t \geqq t_1 > T.$$

But

$$\int_{t_1}^{t_2} e^{-t(s_0+h)}d\alpha(t) = S_{t_1}(t_2,s_0)e^{-ht_2} + h\int_{t_1}^{t_2} S_{t_1}(t,s_0)e^{-ht}dt.$$

Hence

$$\left|\int_{t_1}^{t_2} e^{-t(s_0+h)}d\alpha(t)\right| < \epsilon e^{-(\sigma-\sigma_0)t_2} + \frac{|s-s_0|}{\sigma-\sigma_0}\epsilon\left\{e^{-(\sigma-\sigma_0)t_1} - e^{-(\sigma-\sigma_0)t_2}\right\}.$$

If s is in the region defined in the theorem, this becomes

$$\left| \int_{t_1}^{t_2} e^{-t(s_0+h)} d\alpha(t) \right| < \epsilon + H\epsilon e^{H(\sigma-\sigma_0)} e^{-(\sigma-\sigma_0)t_1}$$
$$< \epsilon + H\epsilon e^{-(\sigma-\sigma_0)(t_1-H)} < (H+1)\epsilon, \quad t_2 > t_1 > T.$$

The theorem is thus established.

COROLLARY 1. *The integral converges uniformly in the angle*

$$|s - s_0| \leqq H(\sigma - \sigma_0), \quad \sigma \geqq \sigma_0.$$

COROLLARY 2. *The integral represents a holomorphic function $f(s)$ in its region of convergence, and in this region*

$$\frac{d^k f(s)}{ds^k} = \int_0^\infty e^{-st}(-t)^k d\alpha(t) \qquad (k = 0,1,2, \cdots).$$

For it is a familiar fact that for any positive numbers a and b the integral

$$\int_a^b e^{-st} d\alpha(t) = e^{-sb}\alpha(b) - e^{-sa}\alpha(a) + s\int_a^b e^{-st}\alpha(t)dt$$

represents an entire function.* Hence to establish the corollary one has only to apply a classical theorem of Weierstrass† regarding uniformly convergent series of analytic functions to the series

$$\sum_{n=0}^{\infty} \int_n^{n+1} e^{-st} d\alpha(t).$$

THEOREM 3. *If the integral* (1.2) *converges absolutely for $s_0 = \sigma_0 + i\tau_0$, it converges uniformly for $\sigma \geqq \sigma_0$.*

The hypothesis implies that for an arbitrary positive ϵ there exists a number T such that

$$\int_{t_1}^{t_2} e^{-\sigma_0 t} du(t) < \epsilon, \quad t_2 > t_1 > T,$$

where $u(t)$ is the total variation of $\alpha(t)$ from zero to t. But‡

$$\left| \int_{t_1}^{t_2} e^{-st} d\alpha(t) \right| \leqq \int_{t_1}^{t_2} e^{-\sigma t} du(t) \leqq \int_{t_1}^{t_2} e^{-\sigma_0 t} du(t) < \epsilon, \quad \sigma \geqq \sigma_0, \quad t_2 > t_1 > T.$$

This proves the theorem.

* Apply, for example, the theorems on p. 282, vol. I of W. F. Osgood, *Funktionentheorie*, 1923, 4th edition.

† W. F. Osgood, loc. cit., p. 303.

‡ For the properties of the Stieltjes integral here employed see, for example, T. H. Hildebrandt, *On integrals related to and extensions of the Lebesgue integrals*, Bulletin of the American Mathematical Society, vol. 24 (1918), p. 180.

We may apply the result of Theorem 2 to the integral (1.4) to establish that it also represents an analytic function within its region of convergence provided that region does not reduce to a linear region.

3. **Abscissa of convergence.** To establish a formula for the abscissa of convergence we shall need two lemmas.

LEMMA 1. *If a real number γ exists for which*

$$|\alpha(t)| < e^{\gamma t}, \quad 0 \leqq t < \infty ,$$

then (1.2) *converges for $\sigma > \gamma$.*

By an integration by parts we have

$$\int_{t_1}^{t_2} e^{-st} d\alpha(t) = e^{-st_2}\alpha(t_2) - e^{-st_1}\alpha(t_1) + s\int_{t_1}^{t_2} e^{-st}\alpha(t)dt,$$

$$\left|\int_{t_1}^{t_2} e^{-st} d\alpha(t)\right| \leqq e^{-\sigma t_2+\gamma t_2} + e^{-\sigma t_1+\gamma t_1} + |s|\int_{t_1}^{t_2} e^{-\sigma t+\gamma t}dt$$

$$\leqq e^{-(\sigma-\gamma)t_2} + e^{-(\sigma-\gamma)t_1} + \frac{|s|}{\sigma-\gamma}\left[e^{-(\sigma-\gamma)t_1} - e^{-(\sigma-\gamma)t_2}\right].$$

If $\sigma > \gamma$ the right-hand side of this inequality may clearly be made as small as desired by taking t_1 and t_2 sufficiently large, so that the lemma is proved.

LEMMA 2. *If* (1.2) *converges for $s = \sigma_0 > 0$, then a constant K exists such that*

$$|\alpha(t)| < Ke^{\sigma_0 t}, \quad 0 \leqq t < \infty .$$

Set

$$U(t) = \int_0^t e^{-\sigma_0 t} d\alpha(t).$$

Then we may write*

$$\alpha(t) = \int_0^t d\alpha(t) = \int_0^t e^{\sigma_0 t}e^{-\sigma_0 t} d\alpha(t) = \int_0^t e^{\sigma_0 t} dU(t).$$

Integrating by parts we have

$$\alpha(t) = U(t)e^{\sigma_0 t} - \sigma_0\int_0^t U(t)e^{\sigma_0 t}dt.$$

Since (1.2) converges for $s = \sigma_0$, a constant K exists such that $|U(t)| < K/2$, $0 \leqq t < \infty$. Hence

$$|\alpha(t)| \leqq \frac{K}{2}e^{\sigma_0 t} + \frac{K}{2}(e^{\sigma_0 t} - 1) < Ke^{\sigma_0 t}.$$

* See T. Carleman, loc. cit., p. 11, Theorem III.

This proves Lemma 2. As an additional result the above method would show that, if (1.2) converges for a negative value of σ, then $\alpha(t)$ would be bounded in the interval $0 \leqq t < \infty$.

THEOREM 4. *The abscissa of convergence of* (1.2), *if it is positive, is given by*

$$\limsup_{t=\infty} \frac{\log |\alpha(t)|}{t} = \sigma_c. \tag{3.1}$$

We prove first that (1.2) converges for $\sigma > \sigma_c$. Let $s_0 = \sigma_0 + i\tau_0$ be an arbitrary point for which $\sigma_0 > \sigma_c$. Let ϵ be chosen so that

$$\sigma_0 > \sigma_c(1+\epsilon), \quad \epsilon > 0.$$

Then

$$\frac{\log |\alpha(t)|}{t} < \sigma_c(1+\epsilon), \quad t > t_0,$$

or

$$|\alpha(t)| < e^{\sigma_c(1+\epsilon)t}.$$

By Lemma 1, (1.2) converges for $\sigma > (1+\epsilon)\sigma_c$, and hence at $s = s_0$.

We now prove that if (1.2) converges for a value $s_0 = \sigma_0 + i\tau_0$, with $\sigma_0 > 0$, then $\sigma_0 \geqq \sigma_c$. For, suppose that $\sigma_0 < \sigma_c$. Choose ϵ so that

$$\sigma_0 < \sigma_0 + \epsilon < \sigma_0 + 2\epsilon < \sigma_c.$$

Since (1.2) converges for $s = \sigma_0$ it also converges for $s = \sigma_0 + \epsilon$, and by Lemma 2,

$$|\alpha(t)| < Ke^{(\sigma_0+\epsilon)t}$$

provided $\sigma_0 + \epsilon > 0$. But by definition of σ_c we have

$$\log |\alpha(t)| > t(\sigma_0 + 2\epsilon)$$

or

$$|\alpha(t)| > e^{t(\sigma_0+2\epsilon)}$$

for certain values of t as large as desired. Since it is impossible to have

$$e^{t(\sigma_0+2\epsilon)} < Ke^{(\sigma_0+\epsilon)t}$$

for large values of t our assumption that $\sigma_0 < \sigma_c$ must have been false. Hence (1.2) converges for $\sigma > \sigma_c$ and diverges for $\sigma < \sigma_c$. The theorem is thus established.

Since in Lemma 1 it was unnecessary to have γ positive, we may be assured that even if σ_c as computed by (3.1) is a negative quantity then (1.2) converges for $\sigma > \sigma_c$. In this case it will not be known, however, that (1.2) does not also converge for $\sigma < \sigma_c$, so that σ_c will not necessarily be the abscissa of convergence.

Finally we note that if $\limsup (\log|\alpha(t)|)/t=\infty$, (1.2) diverges over the entire plane. The proof may easily be supplied.

COROLLARY. *The abscissa of absolute convergence of* (1.2), *if it is positive, is given by*

$$\sigma_a = \limsup_{t=\infty} \frac{\log u(t)}{t},$$

where $u(t)$ is the total variation of $\alpha(t)$ in the interval from zero to t.

It should be pointed out that although formula (3.1) applies only when σ_c is positive, it may always be used indirectly to determine the axis of convergence. For, one has only to displace the origin by a translation to make the formula applicable. It may also be used to determine the region of convergence of an integral (1.4). As an example consider the function

$$\Gamma(s) = \int_{-\infty}^{\infty} e^{-st}e^{-e^{-t}}dt = \int_0^{\infty} e^{-st}e^{-e^{-t}}dt + \int_0^{\infty} e^{st}e^{-e^{t}}dt. \tag{3.2}$$

In the first integral of this sum set $s'=s+1$. Then the function $\alpha(t)$ of formula (3.1) becomes

$$\alpha(t) = \int_0^t e^{-e^{-t}+t}dt$$

and

$$\lim_{t=\infty} \frac{\log \alpha(t)}{t} = 1 = \sigma_c' = \sigma_c + 1.$$

Hence the abscissa of convergence of the first integral is $\sigma_c=0$. In the second integral of (3.2) set $s'=-s+k$, where k is positive but arbitrarily large. Then

$$\alpha(t) = \int_0^t e^{-e^t}e^{kt}dt. \tag{3.3}$$

Since the integral (3.3) is less than e^{-t} for t sufficiently large, it follows that

$$\lim_{t=\infty} \frac{\log \alpha(t)}{t} = 0$$

for all k. That is, the second integral in (3.2) converges for all values of s. Consequently the integral defining $\Gamma(s)$ converges for $\sigma>0$ and diverges for $\sigma<0$.

4. **The uniqueness of the determining function.** We shall show in this section that a given generating function $f(s)$ can not give rise to two deter-

mining functions that have different values at a set of points of positive measure. For, suppose there were two such functions $\alpha_1(t)$ and $\alpha_2(t)$, both vanishing according to our agreement at $t=0$. Then we should have

$$0 = \int_0^\infty e^{-st}d\phi(t), \quad \phi(t) = \alpha_1(t) - \alpha_2(t).$$

This integral must converge for some value $s=s_0$, and hence, by Lemma 2 of §3,

$$|\phi(t)| < Ke^{\sigma_0 t}.$$

Therefore

$$\lim_{t=\infty} e^{-st}\phi(t) = 0, \quad \sigma > \sigma_0,$$

so that

$$\int_0^\infty e^{-st}d\phi(t) = s\int_0^\infty e^{-st}\phi(t)dt = 0, \quad \sigma > \sigma_0.$$

It is now only necessary to apply a result of M. Lerch* to see that $\phi(t)$ is zero except at a set of points of measure zero. Hence $\alpha_1(t)$ and $\alpha_2(t)$ differ at most at a set of points of measure zero, contradicting the assumption. The result is thus established.

5. **Order of $f(s)$ on vertical lines.** As in the case of Dirichlet's series the study of the behavior of $f(\sigma+i\tau)$ as τ becomes infinite (with σ fixed) is of considerable importance. From equation (1.31) we see at once that

$$f(\sigma_1 + i\tau) = O(\tau), \quad \sigma_1 > \sigma_c.$$

For, if σ_0 lies in the interval $\sigma_c<\sigma<\sigma_1$, then

$$\frac{|f(\sigma_1 + i\tau)|}{\tau} \leqq \frac{((\sigma_1 - \sigma_0)^2 + \tau^2)^{1/2}}{|\tau|}\int_0^\infty |S(t,\sigma_0)|\, e^{-(\sigma_1-\sigma_0)t}dt \leqq M, \quad |\tau| \geqq \tau_0.$$

Here M is some constant independent of τ, and τ_0 is any positive constant.

A more general result than this may be obtained, as in the case of Dirichlet's series. We state it in

THEOREM 5. *If the integral* (1.2) *converges for* $s=s_0$, *then*

$$f(\sigma + i\tau) = o(|\tau|)$$

uniformly for $\sigma \geqq \sigma_0+c$, $c>0$.

Let ϵ be an arbitrarily small positive quantity. We wish to show that

* M. Lerch, *Sur un point de la théorie des fonctions génératrices d'Abel*, Acta Mathematica, vol. 27 (1903), p. 339.

there exists a number τ_0 independent of h in the interval $c \leqq h < \infty$ such that

$$\frac{|f(\sigma_0 + h + i\tau)|}{|\tau|} \leqq \epsilon, \quad |\tau| \geqq \tau_0. \tag{5.1}$$

As before we have

$$\int_0^\infty e^{-(\sigma_0+h+i\tau)t}d\alpha(t) = (h + i\tau)\int_0^\infty S(t,\sigma_0)e^{-(h+i\tau)t}dt.$$

We show first that

$$\left|\frac{1}{\tau}\int_a^\infty e^{-(\sigma_0+h+i\tau)t}d\alpha(t)\right| < \epsilon/2 \tag{5.2}$$

for a sufficiently large. We have

$$\int_a^\infty e^{-(\sigma_0+h+i\tau)t}d\alpha(t) = -S(a,\sigma_0)e^{-(h+i\tau)a} + (h + i\tau)\int_a^\infty S(t,\sigma_0)e^{-(h+i\tau)t}dt,$$

$$\left|\frac{1}{\tau}\int_a^\infty e^{-(\sigma_0+h+i\tau)t}d\alpha(t)\right| \leqq \frac{Ae^{-ca}}{\tau_1} + A\left(\frac{1}{c^2} + \frac{1}{\tau_1^2}\right)^{1/2} e^{-ca}, \quad |\tau| \geqq \tau_1, \quad h \geqq c.$$

Since the right-hand member of this inequality is independent of τ and of h for $\tau \geqq \tau_1$, $h \geqq c$, and since it approaches zero as a becomes infinite, we see that (5.2) is established. Likewise we have

$$\left|\frac{1}{\tau}\int_0^a e^{-(\sigma_0+h)t}d\alpha(t)\right| \leqq \frac{1}{|\tau|}\int_0^a e^{-(\sigma_0+h)t}du(t),$$

where $u(t)$ is the total variation of $\alpha(t)$ from zero to t. Since $h \geqq c$ we have a fortiori

$$\left|\frac{1}{\tau}\int_0^a e^{-(\sigma_0+h+i\tau)t}d\alpha(t)\right| \leqq \frac{1}{|\tau|}\int_0^a e^{-(\sigma_0+c)t}du(t).$$

The right-hand side may clearly be made less than $\epsilon/2$ by taking $|\tau|$ sufficiently large, say greater that τ_2. Take τ_0 greater than τ_1 and τ_2. Then (5.1) is established by combining the two inequalities just obtained.

Since this result reduces to a familiar one in the theory of Dirichlet's series when the function $\alpha(t)$ is replaced by a step function, one might be tempted to suppose that all the facts about the order of Dirichlet's series on vertical lines would carry over to the more general generating functions here treated. This is by no means the case. One of the most fundamental results in the theory of Dirichlet's series is that the order of a convergent series is

always positive or zero. That this is no longer necessarily the case for the Stieltjes integral is seen by the example

$$\frac{1}{s} = \int_0^\infty e^{-st}dt.$$

Here the order is clearly negative. Moreover, it is known that for a Dirichlet's series $f(s)$ the limit as τ becomes infinite of $f(\sigma+i\tau)$ can not exist.* This is no longer the case for the functions (1.2) as the above example shows. We shall investigate later the relation between the order of $f(s)$ on vertical lines and the continuity properties of the determining function.

6. **The determination of $\alpha(t)$.** In the theory of Dirichlet's series the formula for the determination of the sum of the first n coefficients is of the utmost importance. In this section we obtain an analogous formula for the Stieltjes integral. The problem here amounts to the solution of the integral equation (1.2) under the assumption that a solution $\alpha(t)$ of bounded variation exists. For the case in which the Stieltjes integral reduces to a Riemann integral this equation is known as Laplace's integral equation. The result to be proved is stated in

THEOREM 6. *If* (1.2) *converges for* $\sigma>\sigma_c$, *and if* c *is a positive constant greater than* σ_c, *then*

$$\frac{\alpha(\omega+0)+\alpha(\omega-0)}{2} = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty}\frac{f(s)}{s}e^{\omega s}ds, \quad \omega>0.$$

Let R be an arbitrary positive constant greater than ω. Then

$$f(s) = \int_0^R e^{-st}d\alpha(t) + \int_R^\infty e^{-st}d\alpha(t),$$

$$\int_{c-i\infty}^{c+i\infty}\frac{e^{\omega s}f(s)}{s}ds = \int_{c-i\infty}^{c+i\infty}\frac{e^{\omega s}ds}{s}\int_0^R e^{-st}d\alpha(t) + \int_{c-i\infty}^{c+i\infty}\frac{e^{\omega s}ds}{s}\int_R^\infty e^{-st}d\alpha(t). \tag{6.1}$$

It will first be shown that the second term on the right-hand side of (6.1) is zero. Set $u=t-R$. Then

$$\int_R^\infty e^{-st}d\alpha(t) = e^{-sR}\int_0^\infty e^{-su}d\bar{\alpha}(u), \quad \bar{\alpha}(u) = \alpha(u+R).$$

Set

* K. Ananda-Rau, *Note on a property of Dirichlet's series*, Proceedings of the London Mathematical Society, (2), vol. 19 (1920–21), p. 114.

$$\phi(s) = \int_0^\infty e^{-su} d\bar{\alpha}(u) \tag{6.2}$$

and consider the integral

$$\int \frac{e^{s(\omega-R)}}{s} \phi(s) ds$$

extended over the rectangle whose vertices are $c-ir, d-ir, d+ir, c+ir$ $(d>c)$. This integral is zero by Cauchy's theorem. The integral

$$\int_{d-ir}^{d+ir} \frac{e^{s(\omega-r)}}{s} \phi(s) ds$$

clearly approaches zero as d becomes infinite, since (6.2) is uniformly convergent in the infinite region $\sigma \geqq c$, $-r \leqq \tau \leqq r$ (Theorem 2). Hence

$$\int_{c-ir}^{c+ir} \frac{e^{s(\omega-R)}}{s} \phi(s) ds = \int_{c-ir}^{\infty-ir} \frac{e^{s(\omega-R)}}{s} \phi(s) ds - \int_{c+ir}^{\infty+ir} \frac{e^{s(\omega-R)}}{s} \phi(s) ds \tag{6.3}$$

provided that these two infinite integrals exist. But by virtue of Theorem 5 we know that to an arbitrary positive ϵ there corresponds a number r_0 such that

$$|\phi(\sigma + i\tau)| < \epsilon |\tau|$$

for $\sigma \geqq c$ and $\tau \geqq r_0$. Hence

$$\begin{aligned} &\left| \int_{c-ir}^{\infty-ir} \frac{e^{(\omega-R)s}}{s} \phi(s) ds \right| \leqq \int_c^\infty \frac{e^{(\omega-R)\sigma}}{r} \epsilon r d\sigma < \frac{\epsilon}{R-\omega}, \\ &\left| \int_{c+ir}^{\infty+ir} \frac{e^{(\omega-R)s}}{s} \phi(s) ds \right| < \frac{\epsilon}{R-\omega}. \end{aligned} \tag{6.4}$$

Consequently both of the infinite integrals in (6.3) converge. Moreover, the inequalities (6.4) show that each of these integrals approaches zero as r becomes infinite. The result stated is thus established.

We turn now to the first integral on the right-hand side of (6.1). It is evidently equal to

$$\lim_{r=\infty} \int_{c-ir}^{c+ir} \frac{e^{\omega s}}{s} ds \int_0^R e^{-st} d\alpha(t) = \lim_{r=\infty} \int_{c-ir}^{c+ir} \frac{e^{\omega s}}{s} \left[e^{-sR} \alpha(R) + s \int_0^R e^{-st} \alpha(t) dt \right] ds.$$

But

$$\int_{c-i\infty}^{c+i\infty} \frac{e^{s(\omega-R)}}{s} ds = 0$$

since* $\omega - R < 0$, and

* See G. H. Hardy and M. Riesz, *The General Theory of Dirichlet's Series*, p. 13.

$$\lim_{r=\infty} \int_{c-ir}^{c+ir} e^{\omega s} ds \int_0^R e^{-st} \alpha(t) dt = \lim_{r=\infty} \int_0^R \alpha(t) dt \int_{c-ir}^{c+ir} e^{s(\omega-t)} ds$$

$$= \lim_{r=\infty} 2i \int_0^R \frac{\alpha(t)}{\omega - t} e^{c(\omega-t)} \sin r(\omega-t) dt .$$

This integral is a Fourier integral,* and the limit is known to be $i\pi[\alpha(\omega+0)+\alpha(\omega-0)]$. The proof of Theorem 6 is thus complete.

7. **Fractional derivatives of the determining function.** If in the integral of Theorem 6 the denominator of the integrand is replaced by $s^{\rho+1}$ where ρ is any real positive number, a new formula of importance may be obtained involving the fractional derivative of order $-\rho$ of the determining function $\alpha(t)$.

THEOREM 7. *If the integral* (1.2) *converges for* $\sigma > \sigma_c$ *and if* c *is a positive constant greater than* σ_c, *then*

$$\frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \frac{f(s)e^{\omega s}}{s^{\rho+1}} ds = D_\omega^{-\rho} \alpha(\omega), \quad \omega > 0, \quad \rho > 0,$$

where $D_\omega^{-\rho}\alpha(\omega)$ *is the fractional derivative of Riemann*:

$$D_\omega^{-\rho}\alpha(\omega) = \frac{1}{\Gamma(\rho)} \int_0^\omega \alpha(t)(\omega - t)^{\rho-1} dt, \quad \rho > 0, \quad \omega > 0.$$

Let R be any real positive number greater than ω. Then

$$f(s) = \int_0^R e^{-st} d\alpha(t) + \int_R^\infty e^{-st} d\alpha(t)$$

and

$$\frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \frac{f(s)e^{\omega s}}{s^{\rho+1}} ds$$

$$= \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \frac{e^{\omega s}}{s^{\rho+1}} ds \int_0^R e^{-st} d\alpha(t) + \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \frac{e^{\omega s}}{s^{\rho+1}} ds \int_R^\infty e^{-st} d\alpha(t).$$

We prove first that the second integral on the right-hand side of this equation is zero. The proof follows closely the lines of that given in §6 for the case in which ρ is zero, so that it will be unnecessary to give the details. One extends the integral

$$\int \phi(s) \frac{e^{s(\omega-R)}}{s^{\rho+1}} ds,$$

where $\phi(s)$ has the same significance as before, over the rectangle defined in §6. By Cauchy's theorem the result is zero. As first d and then r becomes in-

* See, for example, C. Jordan, *Cours d'Analyse*, 1913, 3d edition, vol. 2, p. 277.

finite, the integral extended over three sides of the rectangle approaches zero. The integral over the fourth side approaches

$$\int_{c-i\infty}^{c+i\infty} \frac{e^{\omega s}}{s^{\rho+1}} ds \int_R^\infty e^{-st} d\alpha(t)$$

and must be zero. The inequalities of §6 hold a fortiori when s is replaced by $|s|^{\rho+1}$, since $1/|s|^{\rho+1} < 1/|s|$ when $|s| > 1$.

We turn now to the integral

$$I = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \frac{e^{\omega s}}{s^{\rho+1}} \int_0^R e^{-st} d\alpha(t).$$

If it were permissible to interchange the order of integration we should have

$$I = \int_0^R d\alpha(t) \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \frac{e^{s(\omega-t)}}{s^{\rho+1}} ds.$$

But it is known that

$$\frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \frac{e^{us}}{s^{\rho+1}} ds = \begin{cases} u^\rho/\Gamma(\rho+1), & u \geqq 0, \\ 0 & u \leqq 0, \end{cases}$$

if $\rho > 0$.* Hence it would follow that

$$I = \frac{1}{\Gamma(\rho+1)} \int_0^\omega (\omega - t)^\rho d\alpha(t)$$

$$= \frac{1}{\Gamma(\rho)} \int_0^\omega \alpha(t)(\omega - t)^{\rho-1} dt,$$

and the theorem would be established. It remains then only to justify the interchange in the order of integration. This may be done by showing that the integral

$$\int_{c-i\infty}^{c+i\infty} \frac{e^{s(\omega-t)}}{s^{\rho+1}} ds$$

converges uniformly in the interval $0 \leqq t \leqq R$.† This follows since

* See G. H. Hardy and M. Riesz, loc. cit., p. 50.

† That this is sufficient may be seen by writing

$$f_n(t) = \int_{c-in}^{c+in} \frac{e^{s(\omega-t)}}{s^{\rho+1}} ds$$

and applying Property (7) of Stieltjes integrals as given in the article of T. H. Hildebrandt, loc. cit., p. 181. The admissibility of the change of the order of integration in

$$\int_0^R d\alpha(t) \int_{c-in}^{c+in} \frac{e^{s(\omega-t)}}{s^{\rho+1}} ds$$

can not be held in question.

$$\left| \int_{c-i\infty}^{c+i\infty} \frac{e^{s(\omega-t)}}{s^{\rho+1}} ds \right| \leqq \int_{-\infty}^{\infty} \frac{e^{c(\omega-t)}}{|c+i\tau|^{\rho+1}} d\tau \leqq \int_{-\infty}^{\infty} \frac{e^{c\omega}}{|c+i\tau|^{\rho+1}} d\tau ;$$

the last of these integrals is independent of t and converges.

8. **The order of $f(s)$ on vertical lines as affected by $\alpha(t)$.** In §5 we pointed out that $f(c+i\tau)$ may approach zero as $|\tau|$ becomes infinite. In this section we shall develop certain sufficient conditions imposed on $\alpha(t)$ so that this should be the case. We shall show that $f(c+i\tau)$ may be made to approach zero at least as rapidly as $1/|\tau|^{\rho}$ (ρ being any positive number) by a suitable choice of $\alpha(t)$.

THEOREM 8. *If $\alpha(t)$ has continuous derivatives of orders $1, 2, \cdots, n$ which satisfy the conditions*

$$\alpha^{(k)}(0) = 0 \qquad (k = 0,1, \cdots, n-2), \tag{8.1}$$

$$|\alpha^{(k)}(t)| < Me^{\gamma t} \qquad (0 \leqq t < \infty ;\ k = 0,1,2, \cdots, n) \tag{8.2}$$

for certain constants M and γ independent of k, then

$$f(\sigma_0 + i\tau) = O(|\tau|^{1-n}), \quad \sigma_0 > \gamma.$$

Since $\alpha'(t)$ is continuous, the Stieltjes integral (1.2) reduces to the Riemann integral

$$f(s) = \int_0^{\infty} e^{-st}\alpha'(t)dt,$$

which converges absolutely for $\sigma > \gamma$ by virtue of (8.2). Integrating by parts we have

$$f(s) = -\frac{e^{-st}\alpha'(t)}{s}\Big|_0^{\infty} + \frac{1}{s}\int_0^{\infty} e^{-st}\alpha''(t)dt.$$

By conditions (8.1) and (8.2) the first term of this sum is seen to be zero if $\sigma > \gamma$. Proceeding in this way by successive integration by parts we finally obtain

$$f(s) = \frac{\alpha^{(n-1)}(0)}{s^{n-1}} + \frac{1}{s^{n-1}}\int_0^{\infty} e^{-st}\alpha^{(n)}(t)dt.$$

This integral is absolutely convergent by (8.2), so that

$$f(\sigma_0 + i\tau) = O(|\tau|^{1-n}), \quad \sigma_0 > \gamma.$$

As examples consider the integrals

$$f(s) = \int_0^\infty e^{-st}d(t)^{n-1} = (n-1)!/s^{n-1} = O(|\tau|^{1-n}),$$

$$f(s) = \int_0^\infty e^{-st}d\sin t = s/(s^2+1) = O(|\tau|^{-1}),$$

$$f(s) = \int_0^\infty e^{-st}d(1-\cos t) = +1/(s^2+1) = O(|\tau|^{-2}).$$

This result shows that the order of $f(s)$ is dependent to some extent on the continuity properties of $\alpha(t)$. Conversely, the order of $f(s)$ affects the continuity properties of $\alpha(t)$. In this connection we prove

THEOREM 9. *If* (1.2) *converges for* $\sigma > \sigma_c$, *and if* $f(c+i\tau) = O(|\tau|^\rho)$ *for a positive* $c > \sigma_c$ *and* $\rho < -n$ (n *a positive integer*), *then the function*

$$\phi(\omega) = \frac{\alpha(\omega+0) + \alpha(\omega-0)}{2}$$

is continuous with its first n *derivatives and satisfies the conditions*

$$|\phi^{(k)}(\omega)| < Ke^{\omega c} \qquad (k = 0,1,2,\cdots,n).$$

By Theorem 6 we have

$$\phi(\omega) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \frac{f(s)e^{\omega s}}{s}ds.$$

Then

$$\phi^{(k)}(\omega) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} s^{k-1}f(s)e^{\omega s}ds \qquad (k = 0,1,\cdots,n)$$

provided that these integrals are all uniformly convergent in the arbitrary interval $0 \leqq \omega \leqq \omega_0$. But this fact follows from the inequality

$$\left|\int_{c+ig}^{c+i\infty} (c+i\tau)^{k-1}f(s)e^{\omega s}ds\right| < \int_g^\infty |c+i\tau|^{k-1}M\tau^\rho e^{\omega c}d\tau, \quad g > 0.$$

The dominant integral is independent of ω in the given interval and converges since $\rho < -n$. Since $\phi^{(k)}(\omega)$ is expressed as a uniformly convergent integral, it is continuous. Moreover,

$$|\phi^{(k)}(\omega)| \leqq \frac{1}{2\pi}e^{\omega c}\int_{c-i\infty}^{c+i\infty} |s|^{k-1}|f(s)||ds| = Ke^{\omega c},$$

so that the theorem is proved. The continuity of $\phi(\omega)$ does not of course imply the continuity of $\alpha(\omega)$. But we should not expect the continuity of

$\alpha(t)$ to be completely determined by any property of $f(s)$, since the value of $\alpha(t)$ may be changed at an infinite set of points without altering $f(s)$.

9. The multiplication of generating functions. In this section we shall show that the product of two generating functions is itself a generating function. Let

$$f(s) = \int_0^\infty e^{-st}d\alpha(t), \quad \phi(s) = \int_0^\infty e^{-st}d\beta(t)$$

have abscissas of convergence σ_c and σ_c' respectively. Let $s=\sigma+i\tau$ be a fixed point for which $\sigma>\sigma_c, \sigma>\sigma_c'$, $\sigma>0$. We have seen that

$$f(s) = s\int_0^\infty e^{-st}\alpha(t)dt, \quad \phi(s) = s\int_0^\infty e^{-st}\beta(t)dt \tag{9.1}$$

when these two integrals are absolutely convergent (§3, Lemma 2). Then

$$f(s)\phi(s) = s^2\int_0^\infty e^{-st}\alpha(t)dt\int_0^\infty e^{-sx}\beta(x)dx$$

$$= s^2\int_0^\infty dt\int_0^\infty e^{-s(t+x)}\alpha(t)\beta(x)dx.$$

Set $t+x=y$ and eliminate x:

$$f(s)\phi(s) = s^2\int_0^\infty dt\int_t^\infty e^{-sy}\alpha(t)\beta(y-t)dy. \tag{9.2}$$

If it is permissible to interchange the order of integration, we have

$$f(s)\phi(s) = s^2\int_0^\infty e^{-sy}dy\int_0^y \alpha(t)\beta(y-t)dt.$$

To establish the validity of this interchange we employ a familiar theorem of analysis.* To apply the theorem we introduce a function $K(t, y)$ by the definition

$$K(y,t) = 1, \quad y > t, \quad K(y,t) = 0, \quad y \leqq t.$$

By use of this function, (9.2) may be written

$$f(s)\phi(s) = s^2\int_0^\infty dt\int_0^\infty s^{-sy}\alpha(t)\beta(y-t)K(y,t)dy.$$

* E. W. Hobson, loc. cit., vol. II, p. 347.

Hence we have only to show that

$$\int_0^\infty dt \int_0^\infty e^{-\sigma y} \mid \alpha(t) \mid \mid \beta(y-t) \mid K(y,t)dy$$

converges. This is clearly equal to

$$\int_0^\infty \mid \alpha(t) \mid e^{-\sigma t}dt \int_0^\infty e^{-\sigma x} \mid \beta(x) \mid dx$$

which converges by virtue of the fact that the integrals (9.1) converge absolutely. We have thus established

THEOREM 10. *If the integrals*

$$f(s) = \int_0^\infty e^{-st}d\alpha(t), \quad \phi(s) = \int_0^\infty e^{-st}d\beta(t)$$

converge at a point $s=\sigma+i\tau$ for which $\sigma>0$, then at that point

$$f(s)\phi(s) = s\int_0^\infty e^{-st}d\gamma(t) \tag{9.3}$$

where

$$\gamma(t) = \int_0^t \alpha(x)\beta(t-x)dx = \int_0^t \beta(x)\alpha(t-x)dx. \tag{9.4}$$

It is interesting to see how Cauchy's rule for the multiplication of power series is included in formula (9.3). Let $\alpha(t)$ and $\beta(t)$ be defined as follows:

$$\alpha(0) = 0, \ \alpha(t) = a_0 + a_1 + \cdots + a_{n-1}, \ n-1 < t \leqq n \quad (n = 1,2,3, \cdots),$$

$$\beta(0) = 0, \ \beta(t) = b_0 + b_1 + \cdots + b_{n-1}, \ n-1 < t \leqq n \quad (n = 1,2,3, \cdots).$$

Then

$$f(s) = \sum_{n=0}^\infty a_n e^{-ns}, \quad \phi(s) = \sum_{n=0}^\infty b_n e^{-ns}.$$

To evaluate $\gamma(t)$ suppose that $n-1<t<n$. Then the integral (9.4) may be broken up as follows:

$$\begin{aligned}\gamma(t) &= \sum_{k=0}^{n-1} \int_k^{t-n+k+1} \alpha(x)\beta(t-x)dx + \sum_{k=1}^{n-1} \int_{t-n+k}^k \alpha(x)\beta(t-x)dx \\ &= \sum_{k=0}^{n-1} \int_k^{t-n+k+1} s_k\sigma_{n-1-k}dx + \sum_{k=1}^{n-1} \int_{t-n+k}^k s_{k-1}\sigma_{n-1-k}dx,\end{aligned}$$

where

$$s_k = a_0 + a_1 + \cdots + a_k, \qquad \sigma_k = b_0 + b_1 + \cdots + b_k.$$

Hence

$$\gamma(t) = (t - n + 1)\sum_{k=0}^{n-1} s_k\sigma_{n-1-k} + (n - t)\sum_{k=1}^{n-1} s_{k-1}\sigma_{n-1-k}.$$

To evaluate the integral (9.4) we have

$$s\int_{n-1}^{n} e^{-st}d\gamma(t) = s\int_{n-1}^{n} e^{-st}u_{n-1}dt = [e^{-s(n-1)} - e^{-sn}]u_{n-1},$$

where

$$u_{n-1} = \sum_{k=0}^{n-1} s_k\sigma_{n-1-k} - \sum_{k=1}^{n-1} s_{k-1}\sigma_{n-1-k} = \sum_{k=0}^{n-1} a_k\sigma_{n-k-1}.$$

We thus obtain for the product $f(s)\phi(s)$ the series

$$\sum_{n=1}^{\infty}[e^{-s(n-1)} - e^{-sn}]u_{n-1}. \tag{9.5}$$

If this series is subjected to Abel's transformation, it becomes

$$f(s)\phi(s) = u_0 + \sum_{n=1}^{\infty}(u_n - u_{n-1})e^{-sn} = \sum_{n=0}^{\infty} c_n e^{-sn},$$

where

$$c_n = a_nb_0 + a_{n-1}b_1 + \cdots + a_0b_n.$$

This is the classical Cauchy method of multiplication of power series. The rearrangement of terms of the series (9.5) effected above may be justified by use of Lemma 2, §3.

We turn now to the proof of

THEOREM 11.* *If the integrals*

$$f(s) = \int_{-\infty}^{\infty} e^{-st}\alpha(t)dt, \qquad \phi(s) = \int_{-\infty}^{\infty} e^{-st}\beta(t)dt$$

converge absolutely at a common point $s=\sigma+i\tau$, then

$$f(s)\phi(s) = \int_{-\infty}^{\infty} e^{-st}\gamma(t)dt \tag{9.6}$$

where

$$\gamma(t) = \int_{-\infty}^{\infty} \alpha(x)\beta(t - x)\,dx = \int_{-\infty}^{\infty} \beta(x)\alpha(t - x)dx.$$

* A similar theorem was given by T. Kameda, *Theorie der erzeugenden Funktionen und ihre Anwendung auf die Wahrscheinlichkeitsrechnung*, Proceedings of the Tokyo Mathematical and Physical Society, (2), vol. 8 (1915), pp. 262–295 and 336–360. The conditions imposed, however, are not sufficiently general for our purposes.

The proof of this theorem follows the lines of that of the preceding theorem. We have

$$f(s)\phi(s) = \int_{-\infty}^{\infty} e^{-st}\alpha(t)dt \int_{-\infty}^{\infty} e^{-sx}\beta(x)dx = \int_{-\infty}^{\infty} dt \int_{-\infty}^{\infty} e^{-sy}\alpha(t)\beta(y-t)dy,$$

where $t+x=y$. If now it is permissible to interchange the order of integration, equation (9.6) results. To justify this interchange we employ the same theorem as before and show that

$$\int_{-\infty}^{\infty} dt \int_{-\infty}^{\infty} |e^{-sy}|\,|\alpha(t)|\,|\beta(y-t)|\,dy$$

converges. By setting $y-t=x$ this becomes

$$\int_{-\infty}^{\infty} |\alpha(t)|\,e^{-\sigma t}dt \int_{-\infty}^{\infty} |\beta(x)|\,e^{-\sigma x}dx.$$

This clearly converges since the given integrals converge absolutely.

We show that the familiar formula of Riemann

$$\zeta(s) = \frac{1}{\Gamma(s)} \int_0^{\infty} \frac{x^{s-1}}{e^x - 1}\,dx = \frac{1}{\Gamma(s)} \int_{-\infty}^{\infty} \frac{e^{-st}}{e^{e^{-t}} - 1}dt$$

is only a special case of this theorem. Take the functions $f(s)$ and $\phi(s)$ of Theorem 11 as $\zeta(s)/s$ and $\Gamma(s)$, respectively. Since

$$\zeta(s) = \int_{-\infty}^{\infty} e^{-st}d\alpha(t)$$

with

$$\alpha(t) = 0, \quad t \leqq 0,$$

$$\alpha(t) = n, \quad \log n < t \leqq \log(n+1),$$

then

$$\zeta(s)/s = \int_{-\infty}^{\infty} e^{-st}\alpha(t)dt, \quad \sigma > 1,$$

by Lemma 2, §3. The function $\beta(t)$ of the theorem will be

$$\beta(t) = e^{-e^{-t}}, \quad -\infty < t < \infty.$$

Both $\zeta(s)/s$ and $\Gamma(s)$ are represented by absolutely convergent integrals in the half plane $\sigma>1$, so that the theorem is applicable there. Hence

$$\zeta(s)\Gamma(s) = s \int_{-\infty}^{\infty} e^{-st}\gamma(t)dt, \tag{9.7}$$

where

$$\gamma(t) = \int_0^\infty \alpha(x)e^{-e^{-t+x}}dx. \tag{9.8}$$

Integrating (9.7) by parts we obtain

$$\zeta(s)\Gamma(s) = \int_{-\infty}^{\infty} e^{-st}\gamma'(t)dt,$$

since

$$\lim_{t=\infty} e^{-st}\gamma(t) = 0, \quad \lim e^{-st}\gamma(t) = 0 \quad (\sigma > 1).*$$

But

$$\gamma'(t) = \int_0^\infty \alpha(x)e^{-e^{-t+x}}e^{-t+x}dx. \tag{9.9}$$

The differentiation under the integral sign is justified since the integral (9.9) is uniformly convergent in any finite interval $a < t \leqq b$. For

$$\alpha(x)e^{-e^{-t+x}}e^{-t+x} < e^x e^{-e^{-t+x}}e^{-a+x}, \quad a \leqq t \leqq b,$$

and

$$\int_0^\infty e^{-a+2x}e^{-e^{-b+x}}dx$$

converges. The integral (9.9) may be transformed to a Stieltjes integral, so that

$$\gamma'(t) = \int_0^\infty e^{-e^{-t+x}}d\alpha(x). \tag{9.10}$$

In this transformation we have used the fact that

$$\lim_{x=\infty} \alpha(x)e^{-e^{-t+x}} = 0.$$

Since $\alpha(x)$ is a step function, the Stieltjes integral (9.10) reduces to an infinite series, and

$$\gamma'(t) = \sum_{n=1}^{\infty} e^{-ne^{-t}} = 1/(e^{e^{-t}} - 1).$$

It follows that

$$\zeta(s)\Gamma(s) = \int_{-\infty}^{\infty} e^{-st}/[e^{e^{-t}} - 1]dt.$$

---

* It is easily seen from formula (9.8) that

$$\nu(t) < \int_0^\infty e^x e^{-e^{-t+x}}dx = e^t e^{-e^{-t}},$$

whence the result stated.

It is interesting to note that the functional equation of the Γ-function also results immediately by an application of Theorem 11. For, take

$$\alpha(t) = e^{-e^{-t}}e^{-t}, \qquad \beta(t) = \begin{cases} 1, & t > 0, \\ 0, & t \leqq 0. \end{cases}$$

Then

$$\Gamma(s+1) = \int_{-\infty}^{\infty} e^{-st}\alpha(t)dt, \quad 1/s = \int_{-\infty}^{\infty} e^{-st}\beta(t)dt.$$

By Theorem 11

$$\Gamma(s+1)/s = \int_{-\infty}^{\infty} e^{-st}\gamma(t)dt,$$

where

$$\gamma(t) = \int_{-\infty}^{\infty} \alpha(x)\beta(t-x)dx = \int_{-\infty}^{t} e^{-e^{-x}}e^{-x}dx = e^{-e^{-t}}.$$

Hence

$$\Gamma(s+1)/s = \int_{-\infty}^{\infty} e^{-e^{-t}}e^{-st}dt = \Gamma(s).$$

This the classical functional equation referred to above.

PART II. THE SINGULARITIES OF THE GENERATING FUNCTION AS AFFECTED BY THE DETERMINING FUNCTION

10. **Monotonic determining function.** We propose to study in Part II the analogue of the problem set by Hadamard for power series, to determine the effect of the determining function on the singularities of the generating function. The problem seems scarcely to have been touched for the case of Stieltjes integrals of the form (1.2). The following theorem stated by H. Hamburger* is the only result in this direction that the author has been able to find in the literature:

THEOREM 12. *If $\alpha(t)$ is real and monotonic, then the real point of the axis of convergence of*

$$f(s) = \int_0^{\infty} e^{-st}d\alpha(t) \tag{10.1}$$

is a singular point of $f(s)$.

* Hamburger, *Bemerkungen zu einer Fragestellung des Herrn Pólya*, Mathematische Zeitschrift, vol. 7 (1920), p. 306. A similar theorem was first proved by E. Landau for the corresponding Riemann integral. For the reference see Hamburger's article, footnote 7.

Pincherle has treated the problem for the related integrals $\int_0^{\infty}\phi(t)t^{s-1}dt$ but has obtained no result regarding the addition of singularities. See his memoir *Sur les fonctions déterminantes*, Annales Scientifiques de l'École Normale Supérieure, (3), vol. 22 (1905), p. 9.

First make the change of variable $s'=s-\sigma_c$, σ_c being the abscissa of convergence of (10.1). We have then to consider the integral

$$\int_0^\infty \exp(-s't)\exp(-t\sigma_c)d\alpha(t) = \int_0^\infty \exp(-s't)d\beta(t) \tag{10.2}$$

where

$$\beta(t) = \int_0^t \exp(-t\sigma_c)d\alpha(t),$$

where, for the moment, we write $\exp k$ for e^k because of the complicated exponents. The axis of convergence of (10.2) is the axis of imaginaries, and $\beta(t)$ is clearly a monotonic function. Consequently there is no loss of generality in assuming that $\sigma_c=0$. We may assume further that $\alpha(t)$ is an increasing function. For, if it were decreasing we could replace $\alpha(t)$ by $-\alpha(t)$ and $f(s)$ by $-f(s)$. Suppose that $s=0$ were a regular point of $f(s)$. Then $f(s)$ is analytic in some neighborhood of $s=0$, and the series

$$f(\sigma) = \sum_{n=0}^\infty \frac{(\sigma-1)^n}{n!} f^{(n)}(1)$$

must converge for some value of $\sigma<0$. But

$$f^{(n)}(1) = (-1)^n \int_0^\infty t^n e^{-t}d\alpha(t),$$

so that

$$\begin{aligned} f(\sigma) &= \sum_{n=0}^\infty \frac{(\sigma-1)^n}{n!}\int_0^\infty (-t)^n e^{-t}d\alpha(t) \\ &= \int_0^\infty e^{-t}\sum_{n=0}^\infty \frac{(\sigma-1)^n}{n!}(-t)^n d\alpha(t) \\ &= \int_0^\infty e^{-\sigma t}d\alpha(t), \end{aligned}$$

provided that the integration term by term is permissible. To justify this we note first that the series

$$e^{-t}\sum_{n=0}^\infty \frac{(\sigma-1)^n}{n!}(-t)^n$$

is uniformly convergent in any finite interval $0\leqq t\leqq T$, and that the series

$$\sum_{n=0}^\infty \frac{(\sigma-1)^n}{n!}\int_0^T (-t)^n e^{-t}d\alpha(t)$$

is uniformly convergent in the interval $0 \leqq T < \infty$ since it is dominated* by the series, independent of T,

$$\sum_{n=0}^{\infty} \frac{(1-\sigma)^n}{n!} \int_0^{\infty} t^n e^{-t} d\alpha(t).$$

Hence it follows that (10.1) is convergent for a value of $\sigma < 0$ contrary to the hypothesis that $\sigma_c = 0$. The contradiction shows that $s = 0$ must be a singular point of $f(s)$.

As illustrations of this theorem we recall that the function $\Gamma(s)$ has a singularity at the point $s = 0$, $\zeta(s)$ one at the point $s = 1$.

11. **The addition of singularities.** In this section we seek to generalize a theorem of J. Hadamard concerning the multiplication of the singularities of functions defined by power series. Since a power series in z becomes a Dirichlet series, or an integral of type (1.2), by the transformation $z = e^{-s}$, we ought clearly to be concerned here with the addition of singularities.

Let the integrals

$$(11.1) \qquad f(s) = \int_0^{\infty} e^{-st} d\alpha(t), \quad \phi(s) = \int_0^{\infty} e^{-st} d\beta(t)$$

converge, the first for $\sigma > \sigma_1$, and the second absolutely for $\sigma > \sigma_2$. We suppose that the singularities $\alpha_j = \alpha_j' + \alpha_j'' i$ of $f(s)$, the singularities $\beta_k = \beta_k' + \beta_k'' i$ of $\phi(s)$ and the points γ obtained by adding points α_j to points β_k are all isolated, and that further there exists a number r such that the following conditions hold:

Condition A:

$$\begin{aligned} &|\alpha_k' - \alpha_l'| > r, \quad \alpha_k' \neq \alpha_l', \\ &|\alpha_k'' - \alpha_l''| > r, \quad \alpha_k'' \neq \alpha_l'', \\ &|\beta_k - \beta_l| > r, \\ &|\gamma_k - \gamma_l| > r. \end{aligned}$$

This means that between any two projections of points α on the axes there is a distance greater than r, that the distance between any two points β or between any two points γ is greater than r. Regarding the order of $f(s)$ and $\phi(s)$ we assume further that for an arbitrarily small number η there is a number μ such that

$$f(s) = O(|\tau|^{\mu}), \quad s = \sigma + i\tau,$$

* The fact that $\alpha(t)$ is monotonic enters the proof at this stage.

uniformly for $|s-\alpha_j| \geqq \eta$, and a number ν such that

$$\phi(s) = O(|\tau|^{\nu})$$

uniformly for $|s-\beta_j| \geqq \eta$. Here the numbers μ and ν may be positive, negative, or zero. As examples take $f(s)=1/s$, in which case $f(s)=O(|\tau|^{-1})$ uniformly for $|s| \geqq \eta$; or $f(s)=1/(1-e^{-s})$, in which case $f(s)=O(1)$ uniformly for $|s-2k\pi i| \geqq \eta$, $k=0, \pm 1, \pm 2, \cdots$.

Now let $z=x+iy$ be a point in the common region of convergence of the integrals (11.1) such that $x>\sigma_1+\sigma_2$ and $x>\sigma_2$. Then it is possible to find a positive number c such that $x-c>\sigma_2$ and $c>\sigma_1$. Form the integral

$$F(z) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \frac{f(s)\phi(z-s)}{s^{\rho+1}}ds,$$

where ρ is any number greater than $\mu+\nu$ and greater than μ. We have shown elsewhere* that the integral represents a holomorphic function in the half plane in which the inequalities $x>\sigma_1+\sigma_2$, $x>\sigma_2$ are both satisfied, and that the analytic continuation of this function into the other half plane along lines parallel to the axis of reals is analytic except perhaps at points $\alpha+\beta$ and β.

On the other hand the function $F(z)$ can be expressed as a generating function. To establish this point we make use of a

LEMMA. *Let $\beta(t)$ be a function of bounded variation in any finite interval $0 \leqq t \leqq R$, and denote its total variation in this interval by $u(R)$; let $f(x, t)$ be continuous in the region $0 \leqq x < \infty$, $0 \leqq t < \infty$ and such that the integrals*

$$\int_0^{\infty} |f(x,t)|\,du(t), \qquad \int_0^{\infty} |f(x,t)|\,dx,$$

$$\int_0^{\infty} du(t) \int_0^{\infty} |f(x,t)|\,dx$$

converge. Let the integral $\int_0^{\infty} f(x, t)\,dx$ be continuous for $0 \leqq t < \infty$. Then

$$\int_0^{\infty} d\alpha(t) \int_0^{\infty} f(x,t)dx = \int_0^{\infty} dx \int_0^{\infty} f(x,t)d\alpha(t).$$

The proof follows the lines of that of a corresponding theorem for

* D. V. Widder, *The singularities of a function defined by a Dirichlet series*, American Journal of Mathematics, vol. 49, p. 321. The particular method of representation of the functions $f(s)$ and $\phi(s)$ has no effect on the proof.

Riemann integrals,* and is omitted. We apply the lemma to the interchange of the order of integration in the iterated integral

$$F(z) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \frac{f(s)}{s^{\rho+1}}ds \int_0^\infty e^{-zt}e^{st}d\beta(t). \tag{11.2}$$

Since this integral is improper due to its upper and also to its lower limits of integration, we must apply the lemma twice. We give the details of the application in one case only. The integral

$$\int_0^\infty e^{-t(x-c)} \left| d\beta(t) \right|$$

converges since the integral representing $\phi(s)$ is assumed to converge absolutely for $\sigma > \sigma_2$ and here $x-c>\sigma_2$. Moreover the integral

$$\int_0^\infty \frac{|f(s)|}{|s|^{\rho+1}} e^{ct}d\tau$$

converges since $\rho > \mu$. Finally, the iterated integral

$$\int_0^\infty |d\beta(t)|\, e^{-xt} \int_0^\infty \frac{|f(s)|}{|s|^{\rho+1}} e^{ct}d\tau$$

converges since it may be written as the product of the two foregoing convergent integrals. The change in the order of integration in (11.2) is thus justified, so that we have

$$F(z) = \frac{1}{2\pi i}\int_0^\infty e^{-zt}d\beta(t) \int_{c-i\infty}^{c+i\infty} \frac{f(s)e^{st}}{s^{\rho+1}}ds.$$

But by Theorem 7 it follows that

$$\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \frac{f(s)e^{st}}{s^{\rho+1}}ds = D_t^{-\rho}\alpha(t) \tag{11.3}$$

provided that $\rho > 0$. We may also show that this formula holds if ρ is negative but not an integer. For, in this case the integral

$$\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \frac{f(s)e^{st}}{s^{\rho+k+1}}ds \tag{11.4}$$

converges uniformly in any finite interval $0 \leqq t \leqq R$ for any positive integer k. Chose k so that $\rho+k>0$ but $\rho+k-1<0$. Then by Theorem 7 the integral

* E. W. Hobson, loc. cit., vol. II, p. 347.

is equal to $D_t^{-\rho-k}\alpha(t)$. Morever the integral (11.4) may be differentiated k times under the integral sign since $\rho>\mu$, and the result will be a continuous function. Hence

$$\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty}\frac{f(s)e^{st}}{s^{\rho+1}}ds = \frac{d^k}{dt^k}D_t^{-\rho-k}\alpha(t).$$

But by the definition of fractional derivatives of positive order this is $D_t^{-\rho}\alpha(t)$, so that (11.3) holds for all non-integral values of $\rho>\mu$. We have already seen that if ρ is a negative integer or zero,

$$\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty}\frac{f(s)e^{st}}{s^{\rho+1}}ds = D_t^{-\rho}\psi(t),$$

where

$$\psi(t) = \frac{\alpha(t+0)+\alpha(t-0)}{2}.$$

Consequently

$$(11.5)\qquad F(z) = \int_0^\infty e^{-zt}D_t^{-\rho}\alpha(t)d\beta(t) \qquad (\rho \neq 0, -1, -2, \cdots),$$

$$F(z) = \int_0^\infty e^{-zt}D_t^{-\rho}\psi(t)d\beta(t) \qquad (\rho = 0, -1, -2, \cdots).$$

We sum up the results in

THEOREM 13. *Let*

$$f(s) = \int_0^\infty e^{-st}d\alpha(t)$$

converge for $\sigma>\sigma_1$, *and let*

$$\phi(s) = \int_0^\infty e^{-st}d\beta(t)$$

converge absolutely for $\sigma>\sigma_c$. *Let the singularities* α_k *of* $f(s)$ *and the singularities* β_k *of* $\phi(s)$ *satisfy Conditions* A. *Furthermore, let*

$$f(s) = O(|\tau|^\mu)$$

uniformly for $|s-\alpha_k|\geqq\eta$ *and*

$$\phi(s) = O(|\tau|^\nu)$$

uniformly for $|s-\beta_k|\geqq\eta$ (η *being arbitrarily small*).

Then the function $F(z)$ *defined by*

$$\int_0^\infty e^{-zt} D_t^{-\rho} \alpha(t) d\beta(t) \qquad (\rho \neq 0, -1, \cdots),$$

$$\int_0^\infty e^{-zt} D_t^{-\rho} \frac{\alpha(t+0) + \alpha(t-0)}{2} d\beta(t) \qquad (\rho = 0, -1, \cdots)$$

has singularities at most at the points β and $\alpha+\beta$, where $\rho > \mu+\nu$ and $\rho > \mu$.

COROLLARY 1. *If $\rho = 0$ and if $f(0) = 0$, then the function defined by*

$$\int_0^\infty e^{-zt} \frac{\alpha(t+0) + \alpha(t-0)}{2} d\beta(t)$$

has singularities at most at the points $\alpha+\beta$.

To establish this fact we need to modify only slightly the discussion of the singularities of

$$F(z) = \frac{1}{2\pi i} \int \frac{f(s)\phi(z-s)}{s} ds$$

in the paper already cited. In that paper the origin was excluded from the region under discussion by a loop. In the present case this is no longer necessary since the integrand of the above integral has a removable singularity at the origin. This modification shows that only the points $\alpha+\beta$ (and not the points β) are possible singularities of $F(z)$.

COROLLARY 2. *If ρ is any negative integer, then the function defined by*

$$\int_0^\infty e^{-zt} D_t^{-\rho} \frac{\alpha(t+0) + \alpha(t-0)}{2} d\beta(t)$$

has singularities at most at the points $\alpha+\beta$.

In this case the integrand of the integral defining $F(z)$ is analytic at $s = 0$ unless a point α lies there. In the first case it is unnecessary to exclude the origin from the region in question, and in the second case the points β are included in the set $\alpha+\beta$. The result is consequently as stated in the corollary in either case.

The theorem takes on its simplest form when μ and $\mu+\nu$ are negative and $\alpha(t)$ is continuous, so that ρ may have the value zero. In this case the integral (11.5) takes the simple form

$$\int_0^\infty e^{-zt} \alpha(t) d\beta(t).$$

As an example take $\alpha(t)=t$, $\beta(t)=t$. Then $f(s)=\phi(s)=1/s$. The theorem states that the function defined by

$$\int_0^\infty te^{-zt}dt = 1/z^2$$

has no singularities which are not at $z=0$.

As another example take $\alpha(t)=\sin t$. Then $f(s)=s/(s^2+1)$, $\phi(s)=1/s$. Both functions satisfy the order requirements of the theorem. Consequently the function

$$\int_0^\infty e^{-zt}\sin t\,dt = 1/(z^2+1)$$

can have singularities at most at the points i, $-i$, 0. This example shows that the points $\alpha+\beta$ and β need not be effective singularities. In this case the vanishing of $f(s)$ at the origin explains the disappearance of the points β as singularities. If we interchange the rôles of $f(s)$ and $\phi(s)$ we obtain the result that

$$\int_0^\infty e^{-zt}t\cos t\,dt = (z^2-1)/(z^2+1)^2$$

has singularities at most at the points i and $-i$.

12. **A sufficient condition that a function $f(s)$ should be a generating function.** In the applications which we shall make of Theorem 13 the following result will be useful:

THEOREM 14.* *A sufficient condition that $f(s)$ should be a generating function is that it be analytic at infinity and vanish there.*

Since $f(s)$ is analytic at infinity and vanishes there we can expand it in a power series of the form

$$f(s) = \sum_{n=0}^\infty \frac{a_n n!}{s^{n+1}}.$$

Here the constants a_n must satisfy the condition

$$\limsup_{n=\infty}(|a_n|\,n!)^{1/n} = k, \tag{12.1}$$

where k is a positive number or zero (but is not infinite). Now form the function

$$\sum_{n=0}^\infty a_n t^n. \tag{12.2}$$

* This theorem was first established by Cauchy. See Encyklopädie der Mathematischen Wissenschaften, II, 2, p. 26. The proof there given involves contour integration and is less convenient for our purposes than the proof given here.

On account of equation (12.1) it follows that

$$\lim_{n=\infty} (|a_n|)^{1/n} = 0$$

and the function (12.2) is seen to be an entire function. Now multiply the series (12.2) through by e^{-st} and integrate term by term from zero to infinity. This is permissible provided that

$$a_n \int_0^\infty t^n e^{-st} dt$$

converges (as it clearly does for $\sigma > 0$) and provided that the series

$$\sum_{n=0}^{\infty} a_n \int_0^R e^{-st} t^n dt$$

converges uniformly for $R \geqq R_0$.* To establish this result we have

$$\begin{aligned}\sum_{n=0}^{\infty} \int_0^R a_n e^{-st} t^n dt &\ll \sum_{n=0}^{\infty} |a_n| \int_0^R e^{-\sigma t} t^n dt \\ &\ll \sum_{n=0}^{\infty} |a_n| \int_0^\infty e^{-\sigma t} t^n dt \\ &= \sum_{n=0}^{\infty} \frac{|a_n| n!}{\sigma^{n+1}}.\end{aligned}$$

The dominant series is independent of R, and if $\sigma > k$ it converges by (12.2). Integration term by term is permissible, and we have

$$\int_0^\infty e^{-st} \alpha'(t) dt = \int_0^\infty e^{-st} d\alpha(t) = \sum_{n=0}^{\infty} \frac{\alpha^{(n+1)}(0)}{s^{n+1}} = f(s).$$

The theorem is thus established.

It is not true, conversely, that every generating function is analytic at infinity, as the simplest examples show. The foregoing proof showed that for this to be the case it was necessary that the determining function be entire. But the addition of this condition on the determining function does not make the corresponding generating function analytic at infinity, as the example of the function $\Gamma(s)$ shows. By consideration of the order of the entire function $\alpha(t)$ one does arrive at a necessary and sufficient condition.

---

* We are applying a familiar criterion of Dini. See, for example, T. J. Bromwich, *An Introduction to the Theory of Infinite Series*, 2d edition, 1926, p. 502. The criterion requires further that (12.2) should converge uniformly in an arbitrary interval $0 \leqq t \leqq R$. This is obvious since the series is entire.

THEOREM 15. *A necessary and sufficient condition that the function defined by*

$$f(s) = \int_0^\infty e^{-st} d\alpha(t)$$

be analytic outside the circle $|s| = k$, *have a singularity on the circumference of the circle, and vanish at infinity is that* $\alpha(t)$ *be an entire function of order* unity and of type* k greater than or equal to zero (but not infinite).*

To establish this result we make use of a theorem from the theory of entire functions† which states that a necessary and sufficient condition that $\alpha(t)$ be of order unity and of type k is that

$$\limsup n(|\alpha^{(n)}(0)|/n!)^{1/n} = ke, \quad k \geqq 0. \tag{12.3}$$

Suppose first that the expansion of $f(s)$ in power series about the point at infinity has radius of convergence k. That is, if

$$\alpha'(t) = \sum_{n=0}^{\infty} a_n t^n,$$

then

$$f(s) = \sum_{n=0}^{\infty} \frac{a_n n!}{s^{n+1}} \tag{12.4}$$

has radius of convergence k and

$$\limsup (|a_n| n!)^{1/n} = k, \tag{12.5}$$

whence

$$\limsup n(|a_n|)^{1/n} = ke,$$

since $\lim n/(n!)^{1/n} = e$. But $a_n = \alpha^{(n+1)}(0)/n!$ so that one easily obtains (12.3). Conversely (12.3) implies (12.5), so that an entire function of the kind specified in the theorem leads to a series (12.4) whose radius of convergence is k. The theorem is thus completely established.

As a result of this theorem we see that if

$$f(s) = \int_0^\infty e^{-st} d\alpha(t)$$

is analytic outside the circle $|s| = k > 0$, then‡

* For definitions of order and types of an entire function see L. Bieberbach, *Lehrbuch der Funktionentheorie*, vol. II, p. 228. We are demanding that $\alpha(t)$ be of normal type but not of maximal type.

† L. Bieberbach, loc. cit., p. 231.

‡ L. Bieberbach, loc. cit., p. 228.

$$|\alpha(t)| < e^{(k+\epsilon)|t|}$$

for sufficiently large values of $|t|$ and for all positive numbers ϵ.

13. **Hurwitz's theorem regarding the addition of singularities as a special case of Theorem 13.** It is scarcely necessary to point out that if $\alpha(t)$ and $\beta(t)$ are step functions with points of discontinuity at $t=0, 1, 2, \cdots$, Theorem 13 reduces, by the transformation $z=e^{-s}$, to the familiar theorem of Hadamard regarding the multiplication of singularities of power series. It is, however, more surprising that the classical theorem of Hurwitz on the addition of singularities should also be included as a special case.

Let

$$f(s) = \sum_{n=0}^{\infty} \frac{a_n}{s^{n+1}}, \quad \phi(s) = \sum_{n=0}^{\infty} \frac{b_n}{s^{n+1}},$$

both series being convergent in some neighborhood of the point at infinity. Then, by Theorem 14,

$$f(s) = \int_0^{\infty} e^{-st} d\alpha(t),$$

$$\phi(s) = \int_0^{\infty} e^{-st} d\beta(t)$$

for σ sufficiently large, where

$$\alpha'(t) = \sum_{n=0}^{\infty} \frac{a_n t^n}{n!},$$

$$\beta'(t) = \sum_{n=0}^{\infty} \frac{b_n t^n}{n!}.$$

Let the singularities α of $f(s)$ and those β of $\phi(s)$ satisfy Conditions A. Since $f(s)$ and $\phi(s)$ vanish at infinity we have

$$f(s) = O(|\tau|^{-1}), \quad \phi(s) = O(|\tau|^{-1}).$$

Let us first suppose that $a_0=0$. Then

$$f(s) = \frac{a_0}{s} + \frac{1}{s} \int_0^{\infty} e^{-st} \alpha''(t) dt,$$

$$sf(s) = \int_0^{\infty} e^{-st} d\alpha'(t),$$

$$sf(s) = O(|\tau|^{-1}).$$

Apply Theorem 13 to the functions $sf(s)$ and $\phi(s)$. Clearly we may take $\rho=0$, since $\mu+\nu=-2$ and $\mu=-1$. It follows that

$$\int_0^\infty e^{-zt}\alpha'(t)\beta'(t)dt$$

defines a function with singularities at most at the points $\alpha+\beta$ and β. In the present case we may also be sure that the function has singularities at most at the points $\alpha+\beta$. For, if $s=0$ is not a point α, then $sf(s)$ has a zero at the origin, and we may apply Corollary 1 of Theorem 13. But this integral can be expanded in a power series

$$\alpha'(t)\beta'(t) \doteq \sum_{n=0}^{\infty} \frac{c_n t^n}{n!},$$

where

$$c_n = \sum_{k=0}^{n} \frac{n!}{(n-k)!k!} a_k b_{n-k}.$$

Hence

$$\int_0^\infty \alpha'(t)\beta'(t)e^{-zt}dt = \sum_{n=0}^{\infty} \frac{c_n}{z^{n+1}}. \tag{13.1}$$

This is the result of Hurwitz.*

In the case in which $a_0 \neq 0$, we proceed differently, taking $\rho = -1$. The condition $\rho > \mu+\nu$ of Theorem 13 is still satisfied, but the condition $\rho > \mu$ is now violated, so that special considerations are necessary. For the special case in which $\alpha(t)$ and $\beta(t)$ are entire functions of order unity and of finite type, less stringent conditions on ρ are necessary. The condition $\rho > \mu$ was needed only to show the permissibility of the interchange in the order of integration in

$$I = \int_{c-i\infty}^{c+i\infty} f(s)ds \int_0^\infty e^{st}e^{-zt}d\beta(t).$$

We can verify this directly in the present case. Write

$$f(s) = \psi(s) + \frac{a_0}{s},$$

where

$$\psi(s) = \sum_{n=1}^{\infty} \frac{a_n}{s^{n+1}}.$$

Then

$$I = \int_{c-i\infty}^{c+i\infty} \psi(s)ds \int_0^\infty e^{st}e^{-zt}d\beta(t) + a_0 \int_{c-i\infty}^{c+i\infty} \frac{ds}{s} \int_0^\infty e^{st}e^{-zt}d\beta(t). \tag{13.2}$$

* For a reference see J. Hadamard, *La Série de Taylor et son Prolongement Analytique*, Scientia, 1901, p. 73.

Here $\psi(s)=O(|\tau|^{-2})$. Since $\rho>-2$ we may apply the result obtained in the proof of Theorem 13 (see equation (11.2)) to show that the order of integration may be interchanged in the first of the integrals (13.2). To treat the second of these integrals note that

$$\phi(s) = \frac{b_0}{s} + \frac{1}{s}\int_0^\infty e^{-st}\beta''(t)dt,$$

so that the second integral of (13.2) becomes

$$a_0\int_{c-i\infty}^{c+i\infty}\frac{b_0 ds}{s(z-s)} + a_0\int_{c-i\infty}^{c+i\infty}\frac{ds}{s(z-s)}\int_0^\infty e^{st}e^{-zt}\beta''(t)dt.$$

The first of these integrals is clearly zero, and the order of integration of the second may be interchanged, as we see by again referring to equation (11.2). If in that equation we replace $f(s)$ by the function $1/(z-s)$, for which $\mu=-1$, and take $\rho=0$, then $\rho>\mu$ and the result holds. Hence

$$I = \int_0^\infty e^{-zt}d\beta(t)\int_{c-i\infty}^{c+i\infty}e^{st}\psi(s)ds + a_0\int_0^\infty e^{-zt}d\beta'(t)\int_{c-i\infty}^{c+i\infty}\frac{e^{st}}{s(z-s)}ds$$

$$= \int_0^\infty e^{-zt}[\alpha'(t)-a_0]d\beta(t) + a_0\int_0^\infty [e^{-zt}\beta''(t)(1-e^{tz})/z]dt.$$

By integration by parts we have

$$I = \int_0^\infty e^{-zt}[\alpha'(t)-a_0]d\beta(t) + a_0\int_0^\infty e^{-zt}\beta'(t)dt = \int_0^\infty e^{-zt}\alpha'(t)d\beta(t).$$

This is the integral (13.1) so that the result is the same as before. By Corollary 2 of Theorem 13 we see that the singularities of the function must again be at most at the points $\alpha+\beta$.

14. **Wigert's theorem as a special case of Theorem 15.** As an immediate consequence of Theorem 15 we have the following result, which may be regarded as a generalization of a theorem of Wigert:*

THEOREM 16. *A necessary and sufficient condition that the function defined by*

$$f(s) = \int_0^\infty e^{-st}d\alpha(t)$$

should be analytic in the extended plane except at the origin is that $\alpha(t)$ be entire and satisfy the inequality

$$|\alpha(t)| < e^{\epsilon|t|} \tag{14.1}$$

for an arbitrarily small positive number ϵ.

* Georg Faber, *Über die Fortsetzbarkeit gewisser Taylorscher Reihen*, Mathematische Annalen, vol. 57 (1903), p. 369.

The inequality of the theorem is equivalent to the statement that $\alpha(t)$ is an entire function of order unity and of minimal type. By Theorem 15 the series

$$f(s) = \sum_{n=0}^{\infty} \frac{a_n n!}{s^{n+1}},$$

where

$$\alpha'(t) = \sum_{n=0}^{\infty} a_n t^n,$$

is convergent for $|s|>0$ if and only if the inequality (14.1) is satisfied. The theorem is thus established.

Wigert's theorem appears as a simple corollary to this theorem. For let $\beta(t)$ be defined as follows:

$$\beta(0) = 0, \qquad \beta(t) = n, \qquad n-1 < t \leqq n.$$

Then

$$\phi(s) = \int_0^{\infty} e^{-st} d\beta(t) = \sum_{n=0}^{\infty} e^{-ns} = 1/(1-e^{-s}).$$

Let $f(s)$ be defined as in Theorem 16, and apply Theorem 13 to the functions $f(s)$ and $\phi(s)$. Clearly $\mu=-1$ and $\nu=0$, so that we may take $\rho=0$. Then $\alpha=0$ and $\beta=0, \pm 2\pi i, \pm 4\pi i, \cdots$. Consequently the function defined by

$$\int_0^{\infty} e^{-st}\alpha(t)d\beta(t) = \sum_{n=0}^{\infty} \alpha(n)e^{-ns}$$

has singularities at most at the points $\alpha+\beta=0, \pm 2\pi i, \pm 4\pi i, \cdots$. That is, the function defined by the power series

$$\sum_{n=0}^{\infty} \alpha(n)z^n$$

has no singularities in the extended plane except at the point $z=1$. This is Wigert's theorem.

15. **A generalization of a theorem of Leau.** The following theorem is a generalization of a familiar theorem of Leau regarding power series:*

THEOREM 17. *Let $\alpha(t)$ be a function of bounded variation in every finite interval $0 \leqq t \leqq t_0$, and let $\alpha(t)$ coincide with the function $\psi(t)$ for $t \geqq k$, where $\psi(t)$ is analytic in the region $|t| \geqq k$. Then the function defined by*

$$f(s) = \int_0^{\infty} e^{-st} d\alpha(t) \tag{15.1}$$

* See, for example, G. Faber, loc. cit., p. 371.

is analytic in the entire plane cut along the negative real axis from 0 *to* ∞. *Moreover the points* 0 *and* ∞ *are the only singular points on the cut.*

We first note that the integral (15.1) converges absolutely for $\sigma>0$. For, since $\psi(t)$ is analytic at infinity, there exists a constant M such that

$$|\alpha'(t)| < M, \quad t \geqq k.$$

We break the integral (15.1) into two parts

$$f(s) = \int_0^k e^{-st}d\alpha(t) + \int_k^\infty e^{-st}d\psi(t).$$

Since the function $\psi(t)$ is analytic at infinity, its derivative $\psi'(t)$ is also analytic at infinity and vanishes there. $\psi'(t)$ is therefore a generating function,

$$\psi'(t) = \int_0^\infty e^{-yt}d\beta(y) = \int_0^\infty e^{-yt}\beta'(y)dy.$$

Here $\beta'(y)$ is an entire function of order 1 and of type less than k, since $\psi(t)$ is analytic exterior to and on the boundary of the circle $|t|=k$ (Theorem 15). That is, there exists a constant ϵ so small that

$$|\beta'(y)| < e^{(k-\epsilon)|y|} \tag{15.2}$$

for $|y|$ sufficiently large. Then

$$f(s) = \int_0^k e^{-st}d\alpha(t) + \int_k^\infty e^{-st}dt\int_0^\infty e^{-yt}\beta'(y)dy. \tag{15.3}$$

The integral

$$\int_k^\infty e^{-st}dt\int_0^\infty e^{-yt}|\beta'(y)|\,dy$$

converges by virtue of (15.2). Here it is permissible to interchange the order of integration in the integral in (15.3).* Hence

$$f(s) = \int_0^k e^{-st}d\alpha(t) + \int_0^\infty \frac{\beta'(y)e^{-(s+y)k}}{s+y}dy. \tag{15.4}$$

The first of these integrals is analytic over the entire plane, while the second is analytic in the plane cut along the negative real axis from zero to infinity.†

* E. W. Hobson, loc. cit., vol. II, p. 347.

† This may be seen by changing s to $-s$ and applying a familiar theorem of the theory of functions. See, for example, W. F. Osgood, loc. cit., Hauptsatz, p. 282. The path of integration is there considered to be finite, but only slight modifications are necessary in order to treat the present case. Compare also E. Picard, *Leçons sur quelques Types Simples d'Equations aux Dérivées Partielles*, 1927, p. 65.

It remains only to show that the cut may be altered except at its end points without altering the truth of the above statements. This may be done by showing that the integral

$$\int_C \frac{\beta'(y)e^{-(s+y)k}}{s+y}dy$$

has the same value as the second integral in (15.4), where C is a continuous curve (regular) obtained by slightly displacing any finite set of points on the positive real axis, and where s is any point not between the axis and C. An obvious use of contour integration establishes this point.

To obtain the theorem of Leau as a special case of this theorem we define the function $\phi(s)$ as in the preceding section and apply Theorem 13 to the functions $f(s)$ and $\phi(s)$.

16. **Periodic determining function.** We return now to the proof of

THEOREM 18. *If $\alpha(t)$ is a function of bounded variation of period 2π, then the function defined by*

$$f(s) = \int_0^\infty e^{-st}d\alpha(t) \tag{1.2}$$

has no singularities in the entire plane except perhaps poles of the first order at the points $ni, n = \pm 1, \pm 2, \cdots$.

We note first that the integral (1.2) is convergent for $\sigma > 0$ since a constant M must exist for which $|\alpha(t)| \leqq M$ for all t (Lemma 1, §3). A change in the value of the function $\alpha(t)$ at a set of points forming a denumerable set causes no change in the function $f(s)$. Hence we may suppose, without loss of generality, that at a point of discontinuity ξ of $\alpha(t)$ we have

$$\alpha(\xi) = \frac{\alpha(\xi + 0) + \alpha(\xi - 0)}{2}.$$

Since $\alpha(t)$ is a function of bounded variation it may be expanded in a convergent Fourier's series,

$$\alpha(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos nt + b_n \sin nt, \tag{16.1}$$

$$a_n = \frac{1}{\pi}\int_0^{2\pi} \alpha(t) \cos nt\, dt, \quad b_n = \frac{1}{\pi}\int_0^{2\pi} \alpha(t) \sin nt\, dt. \tag{16.2}$$

We now make use of a theorem from the theory of Fourier's series* regarding

* E. W. Hobson, loc. cit., vol. II, p. 583.

integration term by term. To apply the theorem we need to know that
(a) e^{-st} is of bounded variation over the interval $(0, \infty)$, and
(b) $\int_0^\infty |e^{-st}|dt$ converges.
Both of these facts are evident if $\sigma>0$. Consequently it is permissible to integrate term by term from 0 to ∞. We thus obtain the result

$$\int_0^\infty e^{-st}\alpha(t)dt = \frac{a_0}{2s} + \sum_{n=1}^{\infty} \frac{a_n s + b_n n}{s^2 + n^2}.$$

But

$$f(s) = -\alpha(0) + s\int_0^\infty e^{-st}\alpha(t)dt$$

if $\sigma>0$, so that

$$f(s) = -\alpha(0) + \frac{a_0}{2} + \sum_{n=1}^{\infty} \frac{a_n s^2 + b_n ns}{s^2 + n^2}. \tag{16.3}$$

But we can show that this series defines a function analytic in the entire plane except perhaps at the points ni, $n = \pm 1, \pm 2, \pm 3, \cdots$. For, let m be a positive integer arbitrarily large. The series

$$\sum_{n=m+1}^{\infty} \frac{a_n s^2 + b_n ns}{s^2 + n^2} \tag{16.4}$$

is uniformly convergent in the circle $|s| \leqq m+1/2$, and consequently represents an analytic function there. The uniform convergence of the series may be established as follows. The coefficients a_n and b_n satisfy the inequalities

$$|a_n| < K/n, \quad |b_n| < K/n$$

where K is some constant, since $\alpha(t)$ is a function of bounded variation.* Hence

$$\sum_{n=m+1}^{\infty} \frac{a_n s^2 + b_n ns}{s^2 + n^2} \ll [K(m+\tfrac{1}{2})^2 + K(m+\tfrac{1}{2})] \sum_{n=0}^{\infty} \frac{1}{(\frac{1}{2}+n)^n}, \quad |s| \leqq m + \tfrac{1}{2}.$$

The dominant series is independent of s and converges, so that the desired result is obtained. But $f(s)$ is the sum of the analytic function (16.4) and the rational function

$$-\alpha(0) + \frac{a_0}{2} + \sum_{n=1}^{m} \frac{a_n s^2 + b_n ns}{s^2 + n^2},$$

* E. W. Hobson, loc. cit., vol. II, p. 516.

which has poles of the first order at most at the points $\pm i, \pm 2i, \cdots, \pm mi$ inside the circle $|s| \leqq m+1/2$. Since m was arbitrarily large the theorem is completely established.

As an example let us define $\alpha(t)$ as follows:

$$\alpha(t) = 0, \quad 2n\pi < t < (2n+1)\pi, \quad t = 0,$$

$$\alpha(t) = 1, \quad (2n+1)\pi < t < (2n+2)\pi,$$

$$\alpha(n\pi) = \tfrac{1}{2}, \quad n = 1,2,3, \cdots.$$

Then

$$f(s) = \int_0^\infty e^{-st}d\alpha(t) = \sum_{n=1}^{\infty}(-1)^{n+1}e^{-n\pi s} = 1/(1+e^{\pi s}).$$

By (16.3) we have

$$f(s) = \frac{a_0}{2} + \sum_{n=1}^{\infty}\frac{a_n s^2 + b_n ns}{s^2+n^2},$$

$$a_0 = \frac{1}{\pi}\int_0^{2\pi}\alpha(t)dt = 1,$$

$$a_n = \frac{1}{\pi}\int_0^{2\pi}\alpha(t)\cos nt\,dt = 0, \quad n > 1,$$

$$b_n = \frac{1}{\pi}\int_0^{2\pi}\alpha(t)\sin nt\,dt = \begin{cases} 0, & n \text{ even}, \\ \dfrac{-2}{n\pi}, & n \text{ odd}. \end{cases}$$

Hence

$$\frac{1}{1+e^{\pi s}} = \frac{1}{2} - \frac{2s}{\pi}\sum_{n=1}^{\infty}\frac{1}{s^2+(2n+1)^2}.$$

We have thus obtained the Mittag-Leffler expansion of the function $(1+e^{\pi s})^{-1}$.

17. **The determining function a generating function.** We have seen that the function $\Gamma(s)$ is the sum of two functions

$$\Gamma(s) = \int_0^\infty e^{-st}e^{-e^{-t}}dt + \int_{-\infty}^0 e^{-st}e^{-e^{-t}}dt,$$

the second of which is analytic over the entire plane. Consequently, the first function must have singularities at the points $0, -1, -2, \cdots$. The determining function in this case is

$$e^{-e^{-t}} = \sum_{n=0}^{\infty}\frac{(-1)^n e^{-nt}}{n!},$$

a Dirichlet's series. One is thus led to consider the case in which the determining function $\alpha(t)$ is defined by a general Dirichlet's series

$$(17.1)\quad \alpha(t) = \sum_{n=1}^{\infty} a_n e^{-\lambda_n t},\quad 0 \leqq \lambda_0 < \lambda_1 < \lambda_2 < \cdots,\quad \lim_{n=\infty} \lambda_n = \infty.$$

Since the derivative of a Dirichlet's series is itself a Dirichlet's series, no loss of generality is caused by taking the generating function $f(s)$ in the form

$$f(s) = \int_0^{\infty} e^{-st}\alpha(t)dt.$$

We shall suppose the series (17.1) absolutely convergent for $t \geqq 0$. It will be shown that the series (17.1) may be multiplied by e^{-st} and integrated term by term from 0 to ∞ if $\sigma > 0$, so that

$$(17.2)\qquad f(s) = \sum_{n=0}^{\infty} a_n \int_0^{\infty} e^{-(s+\lambda_n)t}dt = \sum_{n=0}^{\infty} \frac{a_n}{s+\lambda_n},\ \sigma > 0.$$

To justify this step we again apply Dini's criterion, and note first that

$$(17.3)\qquad \int_0^R e^{-st}\alpha(t)dt = \sum_{n=0}^{\infty} a_n \int_0^R e^{-(s-\lambda_n)t}dt,\quad R > 0$$

for every finite R. This follows since a Dirichlet's series converges uniformly in any finite region of the half plane of convergence. It remains only to show that (17.3) converges uniformly in the interval $R \geqq R_0$. But

$$\sum_{n=0}^{\infty} a_n \int_0^R e^{-(s+\lambda_n)t}dt \ll \sum_{n=0}^{\infty} |a_n| \int_0^{\infty} e^{-(\sigma+\lambda_n)t}dt = \sum_{n=0}^{\infty} \frac{|a_n|}{\sigma+\lambda_n}.$$

The dominant series is independent of R and converges for $t=0$, since (17.1) converges absolutely, so that (17.2) is established.

The representation (17.2) of $f(s)$ was established for $\sigma > 0$, but by analytic continuation it holds throughout the region in which the series converges. By virtue of the absolute convergence of the series of coefficients this series converges uniformly and represents an analytic function in any region not including a point $-\lambda_n$. In a neighborhood of a point $-\lambda_n$, $f(s)$ clearly has the form

$$f(s) = \frac{a_n}{s+\lambda_n} + A(s),$$

where $A(s)$ is analytic in the neighborhood of $-\lambda_n$. We have thus proved that $f(s)$ is meromorphic with poles of the first order at most at the points

$-\lambda_0, -\lambda_1, -\lambda_2, \cdots$. The position and character of the singularities of the function $\Gamma(s)$ are thus explained by this result.

Let us consider now the more general case in which $\alpha(t)$ has the form

$$\alpha(t) = \int_0^\infty e^{-ty} d\beta(y),$$

the integral being assumed absolutely convergent for $\sigma \geqq 0$. Then

$$f(s) = \int_0^\infty e^{-st} dt \int_0^\infty e^{-ty} d\beta(y).$$

Appealing once more to the lemma of §11, we see that the order of integration may be interchanged since the integral

$$\int_0^\infty |d\beta(y)| \int_0^\infty e^{-t(s+y)} dt = \int_0^\infty \frac{|d\beta(y)|}{\sigma + y}$$

converges for $\sigma > 0$. Consequently,

$$f(s) = \int_0^\infty \frac{d\beta(y)}{s + y}, \quad \sigma > 0.$$

Although this relation has been established for $\sigma > 0$, we see by analytic continuation that it is valid except on the negative real axis. The function $f(s)$ is thus seen to be analytic* in the entire plane cut from 0 to ∞ along the negative real axis. This cut may be artificial or a genuine boundary of the region of analyticity according to the nature of the function $\beta(y)$. For example, $f(s)$ may reduce to the series

$$f(s) = \sum_{n=0}^{\infty} \frac{\beta_n}{s - a_n}, \tag{17.4}$$

where the points a_n are dense on the negative real axis, and the series

$$\sum_{n=0}^{\infty} |\beta_n|$$

converges absolutely.† In this case Goursat has shown that the line on which the a_n are distributed is a cut for the function $f(s)$. We sum up the results in

---

* T. J. Stieltjes, *Recherches sur les fractions continues*, Annales de la Faculté des Sciences de Toulouse, vol. 8 (1894), p. J 72. Compare also O. Perron, *Die Lehre von den Kettenbrüchen*, 1913, p. 369.

† For a description of the functions of this type see E. Borel, *Leçons sur les Fonctions Monogènes Uniformes d'une Variable Complexe*, 1917, p. 37.

THEOREM 19. *Let the function $f(s)$ be defined by*

$$f(s) = \int_0^\infty e^{-st} d\alpha(t),$$

where $\alpha(t)$ is a generating function

$$\alpha(t) = \int_0^\infty e^{-ty} d\beta(y),$$

absolutely convergent in the interval $0 \leqq t < \infty$. Then $f(s)$ is analytic in the whole plane with the negative real axis removed. This line may or may not be a cut for the function $f(s)$.

18. **The addition of singularities of functions defined by factorial series.** In order to discuss the singularities of functions defined by factorial series we first show the relation of these series to the generating function under discussion. We begin by the proof of

THEOREM 20. *A necessary and sufficient condition that a function $f(s)$ can be developed in a convergent factorial series*

$$f(s) = \sum_{n=0}^{\infty} \frac{a_n n!}{s(s+1)\cdots(s+n)}$$

is that it be a generating function

$$f(s) = \int_0^\infty e^{-st} \alpha(t) dt$$

for which $\alpha(t)$ can be represented by a series of the form

$$\alpha(t) = \sum_{n=0}^{\infty} a_n (e^{-t} - 1)^n \tag{18.1}$$

with

$$|a_n| < n^k \tag{18.2}$$

for some value of k and for n sufficiently large.

We prove first the sufficiency of the condition. By (18.2) we see that

$$\limsup (|a_n|)^{1/n} \leqq 1. \tag{18.3}$$

Hence the radius of convergence of the power series

$$\sum_{n=0}^{\infty} a_n z^n$$

is at least unity, and (18.1) converges for all real non negative values of t. Let us prove that it is permissible to multiply (18.1) by e^{-st} and integrate

term by term from 0 to ∞. We again employ Dini's criterion. It is clear that

$$\int_0^R \sum_{n=0}^{\infty} e^{-st} a_n (e^{-t} - 1)^n dt = \sum_{n=0}^{\infty} a_n \int_0^R e^{-st}(e^{-t}-1)^n dt,$$

for the series (18.1) is uniformly convergent in the arbitrary interval $0 \leqq t \leqq R$. Again

$$\sum_{n=0}^{\infty} a_n \int_0^R e^{-st}(e^{-t} - 1)^n dt \ll \sum_{n=0}^{\infty} \int_0^{\infty} e^{-\sigma t}(1 - e^{-t})^n dt \tag{18.4}$$

$$= \sum_{n=0}^{\infty} \frac{|a_n| n!}{\sigma(\sigma + 1) \cdots (\sigma + n)}.*$$

The dominant series has the same region of convergence as the series

$$\sum_{n=1}^{\infty} \frac{|a_n|}{n^{\sigma}}, \tag{18.5}$$

and the latter is certainly convergent by virtue of (18.2) for $\sigma > k+1$. Series (18.4) is therefore uniformly convergent for $R \geqq R_0$, and term by term integration is justified. Hence

$$\int_0^{\infty} e^{-st}\alpha(t)dt = \sum_{n=0}^{\infty} \frac{a_n n!}{s(s + 1) \cdots (s + n)}, \quad \sigma > k + 1.$$

The sufficiency of the condition is thus established.

For the necessity of the condition, assume that the series

$$\sum_{n=0}^{\infty} \frac{a_n n!}{s(s + 1) \cdots (s + n)} \tag{18.6}$$

converges for σ sufficiently large. Then the series (18.5) converges for some positive value of σ, as $\sigma = k$. Hence its general term approaches zero, and for n sufficiently large,

$$|a_n| / n^k < 1.$$

Define $\alpha(t)$ by the series (18.1). The foregoing proof shows that the corresponding generating function is equal to the series (18.6) for σ sufficiently large and the proof is complete. Other forms of this necessary and sufficient condition have been given by Pincherle† and Nielsen.‡

* For the properties of factorial series here employed, see E. Landau, *Über die Grundlagen der Theorie der Fakultätenreihen*, Sitzungsberichte der mathematisch-physikalischen Klasse der Kgl. Bayerischen Akademie der Wissenschaften zu München, vol. 36, pp. 151–218.

† S. Pincherle, loc. cit., p. 52.

‡ N. Nielsen, *Handbuch der Theorie der Gammafunktion*, 1906, p. 239.

We may now establish a result concerning the addition of singularities of functions defined by factorial series. Let $f(s)$ and $\phi(s)$ be defined by the series

$$f(s) = \sum_{n=0}^{\infty} \frac{a_n n!}{s(s+1)\cdots(s+n)},$$

$$\phi(s) = \sum_{n=0}^{\infty} \frac{b_n n!}{s(s+1)\cdots(s+n)}, \tag{18.7}$$

both convergent for σ sufficiently large. It is known that the series will have a half plane of absolute convergence. By Theorem 20,

$$f(s) = \int_0^{\infty} e^{-st} d\alpha(t),$$

$$\phi(s) = \int_0^{\infty} e^{-st} d\beta(t), \tag{18.8}$$

where

$$\alpha'(t) = \sum_{n=0}^{\infty} a_n (1 - e^{-t})^n, \qquad \beta'(t) = \sum_{n=0}^{\infty} b_n (1 - e^{-t})^n.$$

Let the singularities α of $f(s)$ and those β of $\phi(s)$ satisfy Conditions A, and let these functions satisfy the order conditions of Theorem 13. We are thus in a position to apply that theorem provided that (18.8) converges absolutely. This follows by use of the inequality

$$\int_0^{\infty} |e^{-st}\beta'(t)|\, dt \leqq \int_0^{\infty} e^{-\sigma t} \sum_{n=0}^{\infty} |b_n| (1 - e^{-t})^n dt = \sum_{n=0}^{\infty} \frac{|b_n| n!}{\sigma(\sigma+1)\cdots(\sigma+n)}.$$

This series converges by reason of the absolute convergence of (18.7). We conclude that if $\rho > \mu + \nu$ and $\rho > \mu$, the function

$$F(z) = \int_0^{\infty} e^{-zt} D_t^{-\rho} \alpha(t) d\beta(t) \tag{18.9}$$

has singularities at most at the points $\alpha+\beta$ and β.

There is one case when this result is of particular interest, namely that in which $\rho = -1$. For then (18.9) can be very simply expanded in a factorial series. We have

$$F(z) = \int_0^{\infty} e^{-zt} \alpha'(t)\beta'(t) dt.$$

But

$$\alpha'(t)\beta'(t) = \sum_{n=0}^{\infty} c_n (1 - e^{-t})^n,$$

where

$$c_n = a_0 b_n + a_1 b_{n-1} + \cdots + a_n b_0.$$

Then

$$F(z) = \sum_{n=0}^{\infty} \frac{c_n n!}{z(z+1)\cdots(z+n)}.$$

For a direct application of Theorem 13 with $\rho = -1$ we must have $\mu < -1$. As in the case of Hurwitz's theorem we may take $\mu = -1$ in certain cases. Let us suppose in order to have a situation analogous to that assumed for Hurwitz's theorem, that $f(s)$ and $\phi(s)$ are analytic at infinity, vanishing there. Then μ and ν are at most equal to -1, and if $\rho = -1$, then $\rho > \mu + \nu$. The discussion of the singularities of $F(z)$ given in §11 is consequently valid, and it is only when it comes to expressing $F(z)$ as a generating function that a special discussion is necessary (due to the violation of the condition $\rho > \mu$). For $\rho = -1$

$$F(z) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} f(s)ds \int_0^{\infty} e^{t(s-z)}\beta'(t)dt.$$

Set

$$f(s) = \frac{a_0}{s} + \psi(s),$$

$$\psi(s) = \frac{a_1}{s(s+1)} + \frac{a_2 2!}{s(s+1)(s+2)} + \cdots.$$

Then

$$F(z) = \frac{a_0}{2\pi i}\int_{c-i\infty}^{c+i\infty} \frac{ds}{s}\int_0^{\infty} e^{t(s-z)}\beta'(t)dt + \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty}\psi(s)ds\int_0^{\infty} e^{t(s-z)}\beta'(t)dt. \tag{18.10}$$

Since $\psi(s) = O(|\tau|^{-2})$ the order of integration in the second integral may be interchanged, as was shown in §11. To discuss the first integral integrate (18.8) by parts and obtain

$$\phi(s) = \frac{\beta'(t)e^{-st}}{s}\Big|_0^{\infty} + \frac{1}{s}\int_0^{\infty} e^{-st}\beta''(t)dt.$$

Since $|b_n| < n^k$ for some value of k, we have

$$|\beta'(t)| < \sum_{n=0}^{\infty} n^k(1-e^{-t})^n \ll \sum_{n=0}^{\infty}(n+k)(n+k-1)\cdots(n+1)(1-e^{-t})^n,$$

$$|\beta'(t)| < e^{t(k+1)}.$$

Hence

$$\lim_{t=\infty} \beta'(t)e^{-st} = 0,$$

and

$$\phi(s) = \frac{b_0}{s} + \frac{1}{s}\int_0^{\infty} e^{t(s-z)}\beta''(t)dt.$$

Hence the first integral of (18.10) becomes

$$\frac{1}{2\pi i}a_0\int_{c-i\infty}^{c+i\infty}\frac{b_0}{s(z-s)}ds+\frac{a_0}{2\pi i}\int_{c-i\infty}^{c+i\infty}\frac{ds}{s(z-s)}\int_0^\infty e^{t(s-z)}\beta''(t)\,dt.$$

The first of these integrals is zero, and the order of integration of the second may be interchanged, as we see by again referring to equation (11.2). In that equation replace $f(s)$ by $1/(z-s)$ (for which $\mu=-1$) and take $\rho=0$. Then $\rho>\mu$, and the result holds. Hence

$$F(z)=\frac{1}{2\pi i}\int_0^\infty e^{-zt}\beta'(t)dt\int_{c-i\infty}^{c+i\infty}e^{st}\psi(s)ds+\frac{a_0}{2\pi i}\int_0^\infty e^{-zt}\beta''(t)dt\int_{c-i\infty}^{c+i\infty}\frac{e^{st}}{s(z-s)}ds,$$

$$F(z)=\int_0^\infty e^{-zt}\beta'(t)[\alpha'(t)-a_0]dt+a_0\int_0^\infty e^{-zt}\beta''(t)\left(\frac{1-e^{tz}}{z}\right)dt. \tag{18.11}$$

We have here employed Theorem 6. If we integrate by parts the second integral of (18.11) we have

$$F(z)=\int_0^\infty e^{-zt}[\alpha'(t)-a_0]\beta'(t)dt+a_0\int_0^\infty e^{-zt}\beta'(t)dt=\int_0^\infty e^{-zt}\alpha'(t)\beta'(t)dt.$$

But we have already seen that this is a factorial series,

$$F(z)=\sum_{n=0}^{\infty}\frac{c_n n!}{z(z+1)\cdots(z+n)}.$$

By Corollary 2 of Theorem 13 we see that the function has singularities at most at the points $\alpha+\beta$.

As an example take

$$f(s)=\phi(s)=\frac{1}{s-1}=\sum_{n=0}^{\infty}\frac{n!}{s(s+1)\cdots(s+n)}.$$

All conditions assumed above are satisfied, and since $c_n=n+1$ we conclude that the function defined by the series

$$\sum_{n=0}^{\infty}\frac{(n+1)!}{s(s+1)\cdots(s+n)}$$

has singularities at most at the point $\alpha+\beta=2$. But it is seen by direct computation that the function is $1/(s-2)$, so that the predicted result is verified.

BRYN MAWR COLLEGE,
BRYN MAWR, PA.

A STUDY OF CONTINUOUS CURVES AND THEIR RELATION TO THE JANISZEWSKI-MULLIKIN THEOREM*

BY
LEO ZIPPIN

1. In this paper we treat of continuous curves† in n-dimensional euclidean space; the arguments, excepting the use of inversion,‡ are established in more general space.§ The principal theorems are devoted to the relation of such curves to the Janiszewski-Mullikin Theorem.‖ This, stated generally, is to the effect that two bounded¶ subcontinua of a space, C, neither of which disconnects C, can disconnect C in their sum if and only if their product is not connected.** The theorem is shown to characterise, among bounded cyclicly connected†† continuous curves, the simple closed surface;‡‡ among bounded continuous curves in general, those whose maximal cyclicly connected subsets are simple closed surfaces; among unbounded cyclicly connected continuous curves, the cylinder-trees;§§ and, in general, unbounded

* Various parts of this paper were presented to the Society October 27 and December 1, 1928, and March 30, 1929; it was received by the editors in April, 1929.

† For definitions and theorems, see R. L. Moore, *Report on continuous curves from the viewpoint of analysis situs*, Bulletin of the American Mathematical Society, vol. 29 (1923), pp. 289–302. It is assumed that the reader is familiar with this Report.

‡ See C. Kuratowski, *Sur la méthode d'inversion dans l'analysis situs*, Fundamenta Mathematicae, vol. 4 (1923), pp. 151–163.

§ For a discussion of this space and continuous curves, see H. Hahn, *Mengentheoretische Charakterisierung der stetigen Kurve*, Wiener Sitzungsberichte, vol. 123 (1914), pp. 2433–2489.

‖ See Z. Janiszewski, *Sur les coupures du plan faites par les continus*, Prace Matematyczne-Fisyczne, vol. 26 (1913), p. 48; also, Miss A. Mullikin, *Certain theorems relating to plane connected point sets*, these Transactions, vol. 24 (1922), p. 154. The theorem is readily seen to obtain on the surface of the sphere, from the manner of its proof in the plane.

¶ For the cases that one or both of the continua are unbounded, see B. Knaster and C. Kuratowski, *Sur les continus non-bornés*, Fundamenta Mathematicae, vol. 5 (1924), pp. 35–36.

** When we speak of *our hypothesis*, without further qualification, we shall be understood to refer to this theorem. The Lemma and Theorems 1 and 3 are aside from this hypothesis. Theorem 2 is proved independently of Theorem 5 (of which it is a consequence) and affords the opportunity for introducing methods of proof which are essential to the development of the paper.

†† See G. T. Whyburn, *Cyclicly connected continuous curves*, Proceedings of the National Academy of Sciences, vol. 13 (1927), pp. 31–38. Also, W. L. Ayres, *Continuous curves which are cyclicly connected*, Bulletin de l'Académie Polonaise, 1927, p. 127.

‡‡ A set of points homeomorphic with the surface of the sphere.

§§ I have adopted this term from the analogy with one-dimensional trees, and the relation of this surface to the unbounded cylinder.

continuous curves whose maximal cyclicly connected subsets are cylinder-trees. From the work of G. T. Whyburn and the generalizations of this work by W. L. Ayres, the complete structure of these curves is known. Moreover, the acyclic continuous curves are found to stand in peculiar relation to the Janiszewski-Mullikin Theorem, and it is shown that a continuous curve which may be characterised by this theorem is equivalent in the sense of a Zerlegungsraum* to an acyclic continuous curve of elements, and these are either points of the given curve, or its simple closed subcurves or, exceptionally, the sum of three independent arcs joining two points of the curve.

For his assistance in the solution of the problems of this paper, and for his untiring encouragement, I am greatly indebted to Professor John Robert Kline.

2. We prove the following theorem.

THEOREM 1. *C is a continuous curve, B a closed and totally disconnected subset. There exists in C an acyclic continuous curve which contains B, and whose end points are a subset of B.*†

We assume, at first, that B is bounded; it is immaterial whether C is bounded. Any point x of C is contained in a subcontinuous curve $M(x, \epsilon)$ of C, which is of diameter less than ϵ, a preassigned positive number, and which in some neighborhood U_x^ϵ of x is identical with C.‡ Then x is an interior§ point of $M(x, \epsilon)$. For an arbitrary ϵ',‖ suppose every point b of B covered by continuous curves $M(b, \epsilon')$. Let $\sum U_{b_i}^{\epsilon'}$, where the summation runs from $i=1$ to $i=n_{\epsilon'}$, be a finite covering set of neighborhoods $U_b^{\epsilon'}$ corresponding to these curves, and assemble the curves of $M(b, \epsilon')$ corresponding to the neighborhoods $U_{b_i}^{\epsilon'}$ into maximal connected sets, M_{11}, M_{12}, $\cdots$, M_{1k_1}. Then, since the connected sum of a finite number of continuous curves is a continuous curve, M_{1i} $(i=1, 2, \cdots, k_1)$ is a continuous curve. Let $B_1=\sum_{i=1}^{k_1} M_{1i}$. Every point b of B is an interior point of some subcurve of B_1; we shall say that B is interior to B_1.

Given $B_n=\sum_{i=1}^{k_n} M_{ni}$, we cover B by curves $M(b, \epsilon'_{n+1})$, $\epsilon'_{n+1}<(1/2^n)\epsilon'$,

* See L. Vietoris, *Über stetige Abbildungen einer Kugelfläche*, Koninklijke Akademie van Wetenschappen te Amsterdam, vol. 29 (1926), pp. 443–453.

† See H. M. Gehman, *Concerning acyclic continuous curves*, these Transactions, vol. 29 (1927), p. 566, Theorem 5′.

‡ See H. Hahn, loc. cit., p. 2475, Theorem 21; the condition of boundedness is not necessary to this theorem.

§ A point x is an interior point of a subset X of C, if it is not a limit point of $C-X$. In this case, X is said to cover x in C.

‖ The numbers in this paper are always positive.

such that each curve belongs entirely to $U_{i=1}^{\prime k_n}$ (of the preceding finite covering of neighborhoods) if it has any point in common with $U_{b_i}^{\prime k_n}$. Then, extracting a finite covering of the neighborhoods related to the curves $M(b, \epsilon'_{n+1})$, we assemble the continuous curves corresponding to these neighborhoods into maximal connected sets, $M_{n+1,1}, M_{n+1,2}, \cdots, M_{n+1,k_{n+1}}$; they are continuous curves. Let $B_{n+1}=\sum_{i=1}^{k_{n+1}} M_{n+1,i}$. Then B is interior to B_{n+1}; every curve of B_{n+1} and, therefore, B_{n+1} is interior to B_n.

2.1. There is an arc L'_{11} of C joining a point of M_{11} to a point of $\sum_{i=2}^{k_1} M_{1i}$, and this has a subarc L_{11} from the last point of M_{11} to the first point thereafter on $\sum_{i=2}^{k_1} M_{1i}$; say this is a point of M_{12}. There is an arc L'_{12} joining a point of M_{11} to a point of $\sum_{i=3}^{k_1} M_{1i}$, and a subarc L_{12} from the last point on $M_{11}+L_{11}+M_{12}$ to the first point thereafter on $\sum_{i=3}^{k_1} M_{1i}$; say this is a point of M_{13}. Then $\sum_{i=1}^{3} M_{1i}+\sum_{i=1}^{2} L_{1i}$ is a continuous curve, and every point of $\sum_{i=1}^{2} L_{1i}$ is a cut point of it. Inductively, there is a finite set of arcs $T_1=\sum_{i=1}^{k_1-1} L_{1i}$ such that $T'_1=T_1+B_1$ is a continuous curve, and every point of T_1 is a cut point of that curve. Let $P_1=T_1\times B_1$; P_1 is a finite set of points.

Suppose that we are given T'_n. Let $P_{n1}=P_n\times M_{n1}$, and let $B_{n+1,1}$ be that subset of B_{n+1} which contains all, and consists only, of the curves contained in M_{n1}: $B_{n+1,1}=B_{n+1}\times M_{n1}$. As above, but now in M_{n1}, we find a finite set $T''_{n+1,1}$ of arcs of M_{n1} such that $P_{n1}+T''_{n+1,1}+B_{n+1,1}$ is a subcontinuous curve of M_{n1}, and every point of $T''_{n+1,1}$ (excepting the points P_{n1}) is a cut point. We repeat this for each of the sets M_{ni} $(i=1, 2, \cdots, k_n)$ with respect to $P_{ni}=P_n\times M_{ni}$, and $B_{n+1,i}=B_{n+1}\times M_{ni}$. We have, finally, a set of arcs $T''_{n+1}=\sum_{i=1}^{k_n} T''_{n+1,i}$ such that $T'_{n+1}=T_n+T''_{n+1}+B_{n+1}$ is a subcontinuous curve of T'_n; and every point of $T_{n+1}=T_n+T''_{n+1}$ is a cut point of T'_{n+1}.

We continue this construction for all integral values of n. Then $\overline{T}=\prod_{n+1}^{\infty} T'_n$ is closed and connected. Let $T=\sum_{i=1}^{\infty} T_i$; we shall show that $\overline{T}=T+B$. Since B is interior to every B_n $(n=1,2,\cdots)$, $B\subset\prod_{n=1}^{\infty} B_n\subset\prod_{n=1}^{\infty} T'_n=\overline{T}$. If t is any point of T, there is an n such that t is not a point* of T_n and $t\subset T_{n+1}$. Then $t\subset\prod_{n+1}^{\infty} T_m\subset\prod_{n+1}^{\infty} T'_m$ $(m\geqq n+1)$. Since $T_{n+1}=T_n+T''_{n+1}$, and $T''_{n+1}\subset B_n$, $t\subset B_n\subset\prod_1^n T'_m$ $(m\leqq n)$. Then $t\subset\prod_1^{\infty} \overline{T}'_m=\overline{T}$; and $\overline{T}\subset T+B$. If b' is any point of $\overline{T}$ which does not belong to B, let $r(b', B)=r'>0$.† Find an n such that $(1/2^n)\epsilon'<r'$. Then b' is contained in no $M(b, \epsilon'_{n+1})$, and cannot belong to B_{n+1}. Since $b'\subset\overline{T}\subset T'_{n+1}$, $b'\subset T_{n+1}\subset T$. Then $\overline{T}\subset T+B$, and $T=T+B$.

If t' is any point of T, it is contained in a first T_n; and t' is not a point of B_{n+1} since $T'_n=T_n+B_n$ and B_{n+1} is interior to B_n. Then, in a sufficiently

* To include the case that t is a point of T_1, allow n to take the value zero and define T_0 as the null set.

† $r(X, Y)$ is the distance of the point sets X and Y.

small neighborhood of t', T_{n+1} is identical with $\overline{T}$. Since T_{n+1} is connected im kleinen at all of its points (although not necessarily connected), $\overline{T}$ cannot fail to be connected im kleinen at t', and therefore at any point of T. Since B is totally disconnected, $\overline{T}$ cannot fail to be connected im kleinen at any point.† Then $\overline{T}$ is a continuous curve. If it contains any simple closed curve K, there is an arc k of K no point of which belongs to B, and $k \subset T$. But every point of T is a cut point of $\overline{T}$ since it belongs to some T_n and is a cut point of T_n'; it is clear that such a point separates two subcurves of B_n and therefore at least two points of B, while $\overline{T}$ is a connected subset of T_n' containing every point of B. Then this is not possible.‡ Moreover, it must follow that every end point of $\overline{T}$ is a point of B. Then $\overline{T}$ is the desired acyclic continuous curve.

We have given the necessary condition of Gehman's theorem (he is concerned with the case that B is bounded, and that C is a plane curve); the sufficient condition follows his proof precisely.

2.2. Suppose, now, that B is unbounded; then C is unbounded. Let C^* be the inverse of C with respect to a center of inversion v which is a point of the embedding euclidean space, but not of C.§ Then C^* is a continuous curve, and B^*+v a bounded, closed, and totally disconnected subset. There is in C^* an acyclic continuous curve T' which contains B^*+v, and whose end points are a subset of B^*+v. Then $T'=\sum_i T_i$, where T_i is an acyclic continuous curve (tree), $T_i \times T_j = v$ $(i \neq j)$; and $d(T_m) \leqq e$,‖ where e is preassigned and $m \geqq n_e$.¶ The trees $T_i (i=1, 2, \cdots)$ converge†† to v. It is clear that v cannot be a cut point of C^*; if x^* and y^* are two points of C^*, the corresponding points x and y of C belong to an arc xy of C, and v is not a point of C. Then there is in C^*-v an arc L_2' joining a point of T_1 to a point of T_2. Retain on this arc (which can meet only a finite number of the trees, since these converge to v while L_2' is closed in C^*-v) the subarc from the last point on T_1 to the first point thereafter on T_j $(j>1)$; then, if $j \neq 2$, the further subarc from the last point on T_j to the first point thereafter on T_k $(1 \neq k \neq j)$

† See C. Kuratowski, *Quelques propriétés topologiques de la demi-droite*, Fundamenta Mathematicae, vol. 3 (1922), p. 60, lemme.

‡ See S. Mazurkiewicz, *Un théorème sur les lignes de Jordan*, Fundamenta Mathematicae, vol. 2 (1921), p. 119, lemme.

§ See Kuratowski, §1 loc. cit.; also, in Knaster and Kuratowski, loc. cit., pp. 25–31.

‖ For $d(X)$ read diameter of X.

¶ See K. Menger, *Über reguläre Baumkurven*, Mathematische Annalen, vol. 96 (1926–27), pp 574–575. Also, R. L. Wilder, *Concerning continuous curves*, Fundamenta Mathematicae, vol. 7 (1925), p. 365.

†† We shall say that a sequence of sets $K_1, K_2, K_3, \cdots$, converges to a point x, if x is the unique sequential limit point of any sequence of points $x_1, x_2, x_3, \cdots$, such that x_i is a point of K_i.

and so on inductively until we have a subarc with last point on T_2. If m of the trees have been thus connected, we may suppose the trees renumbered so that these are the first m, and $T_m^* = \sum_1^m T_i + L''$ (this is the set of $m-1$ subarcs of L_2', corresponding to the above process) is a continuous curve, such that v is not a cut point of it, but that every simple closed curve of T_m^* contains v. It is obvious that we can continue in this fashion to join any finite number of the given trees; we wish so to connect all of them, and it is essential that the set of arcs which we thus add to T' converge to v.

We shall show that for any preassigned ϵ, these arcs may be chosen so that not more than a finite number fail to be contained in an ϵ-neighborhood of v. Let $M(v, \epsilon)$ be a subcontinuous curve of C^*, of diameter less than ϵ and in some ϵ'-neighborhood, $U_{v\epsilon'}$, of $v(\epsilon' \leqq \epsilon)$ identical with C^*. If v is a cut point of $M(v, \epsilon)$ there are at most a finite number of distinct components† of $M(v, \epsilon) - v$ (see Lemma); since v is not a cut point of C^* it is readily seen that each component is of diameter, and therefore at upper distance from v, at least ϵ'. Then there is an n such that T_i $(i > n)$ is contained entirely in $U_{v\epsilon'}$, and has no point in any component of $M(v, \epsilon) - v$ if every tree of $\sum_1^n T_i$ has no point in that component. Then if T_k^* $(k \geqq n)$ is constructed, as above, to contain $\sum_1^n T_i$, it is clear that for every tree T_j $(j > k)$ there is an arc L_j such that $L_j \subset (U_{v\epsilon'} - v)$ and joins a point of T_j to a point of T_k^*. Then we are able to define a set of arcs L'' converging to v, such that $T^* = T' + L''$ is connected and closed (since trees and arcs converge to v), that $T^* - v$ is connected, and that every simple closed curve of T^* contains v. It is clear that T^* is a continuous curve since it is connected im kleinen at every point of $T^* - v$, being in a sufficient neighborhood of such points identical with T' plus a finite set of arcs.

Then T, the image of T^*, on C is the desired acyclic continuous curve. It is connected, because $T^* - v$ is connected, and therefore a continuous curve, since T^* is a continuous curve. Also, T is acyclic, for if it contains any simple closed curve K, K corresponds on T^* to a simple closed curve which does not contain v. Moreover, if b is an end point of T its image point b^* is an end point of T^*, since the property of being an interior point of an arc is invariant under inversion (for points other than v), and every end point of T^* is an end point of T' because each arc of L'' has both of its end points on T'. As b^* images b of T, therefore of C, it cannot be the point v; then $b^* \subset B^*$, and consequently $b \subset B$.

3. LEMMA. *A necessary and sufficient condition that a continuum M be a continuous curve is that if L is any bounded subset of M, not more than a finite*

† Maximal connected subsets.

number of the components of $M-L$ are at an upper distance from L exceeding any preassigned ϵ.†

If M fails to be a continuous curve there exists a number ϵ and a point x of M, such that no finite set of subcontinua of M can ϵ-separate x.‡ Since every connected subset of M joining x to a point not in $U_{x\epsilon}$ contains at least one point of the set $N=(\overline{U}_{x(3/4)\epsilon}-U_{x(3/4)\epsilon})$, it is clear that N cannot belong to a finite number of subcontinua of $H=(\overline{U}_{x\epsilon}-U_{x(1/2)\epsilon})$. If, therefore, $L=(\overline{U}_{x\epsilon}-U_{x\epsilon})+(\overline{U}_{x(1/2)\epsilon}-U_{x(1/2)\epsilon})$, the number of components of $H\times(M-L)$ containing points of N is infinite, and these are at an upper distance from L not less than $\frac{1}{4}\epsilon$. This establishes the sufficiency of our condition. These components are also of diameter not less than $\frac{1}{4}\epsilon$ and, by a theorem of Lubben,§ have a countable subsequence with a continuum of condensation M_∞; it is seen that at no point of M_∞ can M be connected im kleinen.‖ It will be apparent that the Moore-Wilder Lemma is implicit in the foregoing.¶

3.1. The condition is necessary. If M is a continuous curve and L is any bounded subset, and if M' is any component of $M-L$, then $r(M', L)=r(M', \overline{L})=0$. Otherwise $M=M'+N'$ (N' is defined as $M-M'$) and is not connected. For if m' is any point of M', it cannot belong to $\overline{L}$ and is an interior point of some component of $M-\overline{L}$;†† and since $M'\times\overline{L}=0$, m' is seen to be an interior point of M'. Then m' is not a limit point of N'. If n' is a point of $\overline{M}'$, n' does not belong to $\overline{L}$ and is an interior point of M'; then n' cannot be a point of N'. Since M is connected, it follows that $r(M', L)=0$. If the upper distance of M' and L is greater than ϵ, from the connectedness of M', there is a point $x\subset M'$ such that $r(x, L)=\epsilon$. If the number of components relative to L‡‡ whose upper distance from L is greater than ϵ is infinite, let (x) be an infinite set of points not more than a finite number of which belong to a single component of $M-L$, and such that each is at a distance from L equal to ϵ. Since L is bounded, $L+(x)$ is bounded, and (x) is bounded.§§ Let x' be any limit point of (x), and (x_i) a subsequence of (x)

† The Lemma will be found to resemble in its details a number of known results. The author prefers to regard it as an opportunity for justifying, in a measure, his use of theorems for which he contemplates a more general space (see §1 third note) than that for which their proof is explicit.

‡ See P. Urysohn, *Über im kleinen zusammenhängende Kontinua*, Mathematische Annalen, vol. 98 (1927), p. 297, Theorem 1.

§ See R. G. Lubben, *Concerning limiting sets in abstract spaces*, these Transactions, vol. 30 (1928), p. 675, Theorem 5.

‖ See §2.1, third note. ¶ See the last paragraph of §2.1.

†† See C. Kuratowski, *Une définition topologique de la ligne de Jordan*, Fundamenta Mathematicae, vol. 1 (1920), p. 40.

‡‡ A component relative to X of M is a component of $M-X$.

§§ This is the only purpose in our restriction on L. Without it the Lemma, as stated, is untrue but is readily modified.

of which x' is the sequential limit. It is clear that $r(x', L) = \epsilon$, and x', and therefore all but a finite number of the points of (x_i), belong to a single component of $M - L$ contrary to our choice of the set (x).

3.2. **Remark.** If M is a continuous curve and L is a closed and bounded subset, the complement of L in M is open and is the sum of a countable set of components; each component has at least one limit point on L, and is connected im kleinen.† Then L plus any number of the components, entire, relative to L is closed and connected, although not necessarily bounded. Also, if L is a continuous curve, then $L^* = L + \sum' M_i$ (where M_i is a component relative to L, and the prime indicates that the summation does not include all of the components of $M - L$) is a continuous curve. For L^* is clearly connected im kleinen at all points of $\sum' M_i$. If p is any point of L, for a preassigned ϵ there is an ϵ' such that any point of $L \times (U_{p\epsilon'})$ can be joined to p by an arc of L of diameter less than ϵ. Also, there is an ϵ'' such that any point of $M \times (U_{p\epsilon''})$ can be joined to p by an arc of diameter less than ϵ'. It is readily seen that any point of $L^* \times (U_{p\epsilon''})$ can be joined to p by an arc of L^* of diameter less than ϵ.

Suppose, finally, that K is a simple closed curve of M. If x of K is a cut point of M, $K - x$ is contained in a single component of $M - x$, and there is another component M_x relative to x which has no limit point on $K - x$. Then M_x is seen to be a component also of $M - K$. The number of these is countable. If y of K is a cut point of M distinct from x, there is a component M_y relative to y, and relative also to K, which is distinct from M_x. Then the number of cut points of M on K is countable.‡

4. We prove the following theorem:

THEOREM 2. *C is a continuous curve of dimension one, containing at least one simple closed curve K. Then C cannot satisfy the Janiszewski-Mullikin Theorem.*§

Let x be a point of K such that $C - x$ is connected. Form an ϵ-separation of C at x, $\epsilon < d(K)$.‖ Then $C = A + B + D$, $A \times B = B \times D = \overline{A} \times D = A \times \overline{D} = 0$, $A \supset x$, $A + B \subset U_{x\epsilon}$, and B is closed and totally disconnected. Then there is in C an acyclic continuous curve T whose end points are a subset of B and which contains B: see Theorem 1.

† These are ready consequences of the proof of the Lemma. Compare, moreover, C. Kuratowski, §3.1 first note, and *Sur les continus de Jordan et le théorème de M. Brouwer*, Fundamenta Mathematicae, vol. 8 (1926), pp. 136–149.

‡ Compare the lemma of Mazurkiewicz, §2.1 fourth note.

§ See §1, and the fourth note thereto.

‖ See in P. Urysohn, *Sur les multiplicités Cantoriennes*, Fundamenta Mathematicae, vol. 7 (1925), pp. 65–72. Also for all other references to dimension.

If x belongs to T it is of finite order on T. Otherwise x is a limit point of end points of T,† which is not possible since these belong to B. Likewise x is not a limit point of branch points of T. On each branch T_j of T of which x is the foot, choose a point p_j such that the arc xp_j contains no point of B and no branch point of T. Omit from T the arcs $(xp_j - p_j)$. In $C-x$, join the points (p_j) by a set of arcs (p_1p_k), where p_1 is a fixed one of the points and p_k the remaining points in succession. The set $T-(xp_j-p_j)+(p_1p_j)$ is a continuous curve which contains the set B and is free of x. It has a subacyclic continuous curve which contains B; call this T'.

Let M_x be the maximal connected subset of $C-T'$ which contains x, and $B'=T'\times\overline{M}_x$. Let T_x be a subcontinuum of T' which is irreducibly connected about B'.‡ Then T_x is an acyclic continuous curve.§ Since the end points of T_x are non-cut points of T_x, it follows that the end points of T_x are a subset of B'.‖ Since $M_x \subset C-T'$ and $C-T' \subset C-B=A+D$, while $M_x \times A \supset x$, it follows that $M_x \subset A$. Then $\overline{M}_x \subset \overline{A}=A+B$. Let y be a point of $K\times D$; by our choice of ϵ at least one such point exists. Then $y\times\overline{M}_x \subset y\times\overline{A}=0$, and $r(y, \overline{M}_x)=\delta_y>0$. There are two points a and b of $K\times D$ such that the arc ayb of $K\times D$ satisfies the relation $r(ayb, \overline{M}_x)\geqq\frac{1}{2}\delta_y$.

We shall suppose first that $ayb \subset T_x$. If z is a point of ayb which is a branch point of T_x, the corresponding branch or branches of T_x contain at least one end point of T_x, and therefore at least one point of B'. But $B' \subset \overline{M}_x$. Then $r(z, B')\geqq r(ayb, \overline{M}_x)\geqq\frac{1}{2}\delta_y$, and these branches are at least of diameter $\frac{1}{2}\delta_y$. But the number of such branches, and therefore the number of corresponding branch points z of ayb, is finite. Then there is an arc $a''b''$ of ayb which contains no branch point of T_x. This has a subarc $a'b'$, where a' and b' are non-cut points of C.

4.1. By the Janiszewski-Mullikin Theorem, $C-(a'+b')$¶ is connected. Then there is an arc of $C-(a'+b')$ joining an interior point o' of $a'b'$ to the point x. This has a subarc px, where p is the last point on $[a'b']$, in order from o' to x. Since p is of order two on T_x,†† in some neighborhood of p no point of px belongs to T_x. Since $p \subset D$ and $x \subset M_x$, while $D\times M_x=0$, there is on the arc px, in order from p, a first point q of $\overline{M}_x$; this is not a point of M_x.

† See §2.2 and the third note thereto.

‡ See W. A. Wilson, *On the oscillation of a continuum at a point*, these Transactions, vol. 27 (1925), §6, p. 433.

§ See S. Mazurkiewicz, loc. cit., p. 123, lemme.

‖ See H. M. Gehman, *Irreducible continuous curves*, American Journal of Mathematics, vol. 49 (1927), p. 190, Theorem 3.

¶ See Remark at conclusion of this theorem.

†† The arc $a'p$ and pb' belong to T_x, while no point of this arc is on T_x a point of order three or more (branch point).

Then $q \subset \overline{M}_x - M_x \subset T'$, since M_x is maximally connected in $C - T'$. Then $q \subset \overline{M}_x \times T' = B' \subset T_x$. Also, $pq \times M_x = 0$. Therefore $C - T_x$ is not connected. And if, contrary to our first supposition, some point y' of ayb does not belong to T_x, then y' belongs to an arc $py'q$ of K whose end points only belong to T_x, and $py'q \times M_x = 0$. In either case, $C - T_x$ is not connected.

The points p and q belonging to T_x, there is an arc pq of T_x. The two arcs pq having only their end points in common, form a simple closed curve. On the arc pq of T_x at most a countable set of points can be branch points of T_x, and at most a countable number cut points of C. Let t be an interior point of this arc which is neither a cut point of C nor a branch point of T_x. Then t separates p and q on T_x. Also, $T_x = I_1 + I_2$; I_1 and I_2 are trees, $I_1 \times I_2 = t$, $I_1 \supset p$, $I_2 \supset q$.

4.2. Suppose $C - I_1$ is not connected; then $C - I_1 = M_{11} + M_{12} + \cdots$, where M_{1i} is a component of $C - I_1$. Let $M_{11} \supset I_2 - t$. Then $M_{11} \supset q$, and therefore $M_{11} \supset [pq]$.† Also, I_1 has at least one end point of T_x, and this is a point of B' and therefore a limit point of M_x. Then $M_{11} \supset M_x$. Form $I_1^* = I_1 + M_{12} + M_{13} + \cdots$; $C - I_1^* = M_{11}$, and is connected. In the same manner form I_2^*; $C - I_2^* = M_{21}$, and $M_{21} \supset (I_1 - t) + [pq] + M_x$. Then I_1^* and I_2^* are two continua neither of which disconnects C.‡

The set $I_1^* + I_2^*$ is a continuum, since $I_1^* \times I_2^* \supset t$. Moreover, $C - (I_1^* + I_2^*) \subset C - (I_1 + I_2) = C - T_x$. But $C - (I_1^* + I_2^*)$ contains $[pq]$ and M_x; then it is not connected. Suppose $I_1^* \times I_2^* \supset t'$ distinct from t. Then t' is not a point of M_{11} or of M_{12} and therefore it cannot belong to I_2 or to I_1. But there is in $C - t$ an arc $t's'$ of which s', distinct from t, is the only point on T_x. If $s' \subset I_1$, then $s't' \subset M_{12}$; if $s' \subset I_2$, $s't' \subset M_{11}$. Therefore $I_1^* \times I_2^* = t$, and I_1^* and I_2^* are two continua neither of which disconnects C, whose sum disconnects C, and whose product is connected. This contradicts the Janiszewski-Mullikin Theorem.

Remark. If it is desired to restrict our hypothesis so that it does not comprehend the case that the continua are points, it is possible to reduce this case to the restricted hypothesis, for continuous curves. If x and y are two non-cut points of a bounded continuous curve C, such that their sum disconnects C, there are two arcs xoy and $xo'y$ in C which have only their end points in common so that they form a simple closed curve K, and which belong to no connected subset of $C - (xy)$. Let I_1 and I_2 be two arcs of K which contain x and y respectively, and are without common point.

---

† The symbol $[pq]$ denotes $pq - p - q$. We understand here that arc pq which has no point in common with M_x and only its end points on T_x.

‡ See Remark to Lemma. If $C - I_j$ is connected, $I_j^* = I_j (j = 1, 2)$.

Forming I_1^* and I_2^* as before, we shall find that they contradict our restricted hypothesis.

5. We prove the following theorem.

THEOREM 3. *C is a one-dimensional continuous curve, bounded or unbounded, not a simple closed curve. Then C is disconnected by some acyclic subcontinuous curve.*

If C is itself acyclic, there is an arc which disconnects it. If C contains a simple closed curve K, precisely as before we construct the acyclic continuous curve T_x. Rehearsing the argument of the previous theorem, we find that either T_x disconnects C, or there are two non-cut points a' and b' on K whose sum disconnects C. Since C is not the simple closed curve K, one of the arcs $a'b'$ of K disconnects C.

Although for any proper subcontinuous curve of the plane, and whether it is one- or two-dimensional, the acyclic continuous curve of this theorem may be replaced by an arc, I do not know whether this may be done in general.

6. We prove the following theorem:

THEOREM 4. *C is a cyclicly connected† continuous curve satisfying the Janiszewski-Mullikin Theorem. Then C is a simple closed surface.‡*

No arc of C disconnects C. Suppose that L is an arc xy of C such that $C-L=N_1+N_2+N_3+\cdots$, in maximally connected sets, at least two being distinct. Since no point of C is a cut point of C, each set N_i has at least two

† See G. T. Whyburn, and W. L. Ayres, §1 seventh note. Any two points of a cyclicly connected continuous curve C belong to a simple closed subcurve of C; for this it is necessary and sufficient that C have no cut point. The author has devised a generalization of this which is valid in compact metric space; the proof of W. L. Ayres has not yet been published.

‡ After this paper was in the hands of the editors, there was received in this country volume 13 of the Fundamenta Mathematicae. In this volume, Casimir Kuratowski in an article entitled *Une charactérisation topologique de la surface de la sphère* gives a most interesting demonstration of the above theorem. It results from the work of Kuratowski that if the Janiszewski-Mullikin Theorem be expressed as two theorems (as is done in the paper of Janiszewski, loc. cit.), then these theorems are equivalent for bounded continuous curves, with a consequent material reduction of our hypothesis. It had seemed, therefore, as if we might well omit our proof even though it little resembles that of Kuratowski and is related to a very different body of supporting theorems. However, it will be evident that to our proof of Theorem 6 the arguments of Theorem 4 are absolutely essential, and that we should be obliged to reproduce the demonstration of Theorem 4 almost in its entirety as argument to Theorem 6; for this reason it is retained. And although the methods of Kuratowski would permit a considerable simplification of our work, it has seemed proper to acknowledge that fact and to leave the manuscript in its original form.

In view of the equivalence referred to above, only one case needs to be considered in Theorem 5. For unbounded continuous curves the equivalence does not obtain, but it is still true that Theorem A implies Theorem B (see Janiszewski, loc. cit.). The Theorems 5 and 6 are not part of Kuratowski's paper.

distinct limit points on L. Let x' and y' be the first and last points on L, in the order $xx'y'y$, which are limit points of N_1. Since $C-(x'+y')$ is connected it contains an arc pz, where p is any point of $C-(L+N_1)$ and $z \subset [x'y']$. This has a subarc pz', z' being the first point from p on $x'y'$. No point of pz' belongs to N_1. Otherwise, since p is not in N_1, there is a point n' on pz' such that $n' \subset \overline{N}_1$, and is a limit point of points not in N_1. Then, $\overline{N}_1-N_1$ being contained in $x'y'$, it follows that n' is a point of $[x'y']$ and contradicts our choice of z'. The arc $z'p$ has a subarc $z'p'$ of which z' is the only point on L. Then $z'p'-z'$ belongs to a single component relative to L, not N_1; say this is N_2. Then z' is a limit point of N_2. Let z'' be any other point of L which is a limit point of N_2. The arcs $x'y'$ and $z'z''$ have a common subarc; let t be an interior point of this arc. Then t separates x' and y' and also z' and z''. Calling xt of L the set I_1, and yt the set I_2, we form the sets I_1^* and I_2^*, as in Theorem 2, and obtain the same contradiction of our hypothesis.

It readily follows that if K is any simple closed curve of C, $C-K$ is not connected.† Then $C-K=M_1+M_2+M_3\cdots$, in maximally connected sets, at least two being distinct. Each set M_i has at least two distinct limit points on K. Suppose that the point y of K fails to be a limit point of M_1. Then, from y in either direction on K, there is a first point which is a limit point of M_1; let these be x and z, and y' any point of that arc xz on K which does not contain y. Then $\overline{M}_1-M_1 \subset xy'z$. Therefore $C-xy'z=M_1+([xyz]+\sum_{i>1} M_i)$, and is not connected; for M_1 consists of interior points, and no point of the bracket can belong to $\overline{M}_1$. But no arc of C can disconnect it. Therefore every point of K is a limit point of each of the sets M_i; symbolically $\overline{M}_i-M_i=K$. We wish to show that there are not more than two K-domains.‡

6.1. We adopt the following notation: $U_{p\epsilon}$ is any ϵ-neighborhood of a given point p, and the corresponding $U_{p\delta}$ is any δ-neighborhood where δ is so chosen that any point of C whose distance from p is not greater than δ can be joined to p by an arc of C of diameter less than ϵ; an arc, therefore, which is contained in $U_{p\epsilon}$. Moreover, if q is any point of $U_{p\delta}$ and we construct the arc pq, we shall understand that $pq \subset U_{p\epsilon}$.

On K choose six points in the order $a'b'oc'd'o'$. Find $U_{a'\epsilon}$ and $U_{d'\epsilon}$ such that $U_{a'\epsilon}\times(b'oc'd'o')=U_{d'\epsilon}\times(o'a'b'oc')=0$. Let q be any point of $U_{a'\delta}\times M_1$ and q' any point of $U_{d'\delta}\times M_1$. Construct the arcs $a'q$ and $d'q'$, and in M_1 any arc qq'. The sum of the three arcs is a continuous curve and

† Compare C. Kuratowski, §3.2 loc. cit., p. 145, Theorem 2.

‡ A K-domain is a component relative to K.

contains an arc $a'd'$. This has a subarc amd, where $m \subset M_1$, such that $[amd] \subset M_1$, $a \subset K \times U_{a'\epsilon}$, and $d \subset K \times U_{d'\epsilon}$; then we have on K the order $ab'oc'do'$. Similarly we construct the arc $bm'c$, such that $[bm'c] \subset M_2$, $b+c \subset K$, and we have the order $abocdo'$. Then $amd+dc+cm'b+ba$, where dc and ba are the arcs of K which contain neither o nor o', is a simple closed curve K''. If we assume that $C-K$ contains at least three K-domains, M_1, M_2, and M_3, we shall show that $C-K''$ is connected.

Suppose, otherwise, that $C-K''=M+N$, $M \times \overline{N} = \overline{M} \times N = 0$. Since K'' has no point in common with M_3, M_3 is connected in $C-K''$ and belongs entirely to M or entirely to N; say $M_3 \subset M$. Then, $\overline{M}_3 \subset \overline{M} \subset M+K''$. Since $[ao'd]+[boc] \subset \overline{M}_3 \subset M+K''$, and $([ao'd]+[boc]) \times K''=0$, it follows that $[ao'd]+[boc] \subset M$; $[ao'd]$ is that arc of K which does not contain o, and $[boc]$ the arc of K which does not contain o'. Therefore $M \supset \sum_{i \geq 3} M_i$ (if this is not vacuous), since each $M_i (i \geqq 3)$ is connected in $C-K''$, while $\overline{M}_i \supset [ao'd]+[boc]$ $(i \geqq 1)$. There remains of $C-K''$ the set $M_1-[amd]$ (the case for $M_2-[bm'c]$ is similar). Since o' is a limit point of M_1 and therefore of $M_1-[amd]$, if $M_1-[amd]$ is connected it belongs to M. Let M_{11} be any maximal connected subset. If any point of $[ao'd]+[boc]$ is a limit point of M_{11}, $M_{11} \subset M$. Then every point of K which is a limit point of M_{11} belongs to one of the arcs ab or cd, and therefore to $bamdc$. It is seen that M_{11} is maximally connected in $C-(K+amd)$ so that it consists of interior points. On the other hand, every limit point of M_{11} which belongs to M_1 is contained in $M_{11}+amd$, while every limit point of M_{11} which does not belong to M_1 is contained in K, and therefore in $ab+cd$. Then $\overline{M}_{11}-M_{11} \subset bamdc$, and $C-bamdc=M_{11}+H$ (it is sufficient to regard H as defined by this relation) and is not connected. This is not possible. Then $M \supset M_{11}$, and likewise every maximal connected subset of $M_1-[amd]$ so that $M \supset M_1-[amd]$. Similarly $M \supset M_2-[bm'c]$, and N is vacuous. Then $C-K''$ is connected; since this is impossible, it follows that there are precisely two K-domains, which we shall call D_1 and D_2. Then $\overline{D}_i-D_i=K$ $(i=1, 2)$; in the sequel, $D_i=M_i$ $(i=1, 2)$.

The arc $[amb]$ disconnects D_1. For $J_1=amboa$ disconnects C. But $C-J_1=(D_1-[amb])+[ao'b]+D_2$, and if $D_1-[amb]$ is connected, then $C-J_1$ is connected. Moreover, $C-J_1=D_{11}+D_{21}$; say that D_{21} contains a point d' of D_2. Any other point d'' of D_2 can be joined to d' by an arc $d'd''$ of D_2; $d'd''$ is free of points of D_1+K, and therefore of points of J_1. Then $d'd'' \subset D_{21}$, and consequently $D_{21} \supset D_2$. Then $D_{21} \supset [ao'b]$. Therefore D_{11} is a maximal connected subset of $D_1-[amb]$. Similarly, if $J_2=amb+[ao'b]$, $C-J_2=D_{12}+D_{22}$, and D_{12} is a maximal connected subset of $D_1-[amb]$. Then $D_1-[amb]=D_{11}+D_{12}+\cdots$, and we shall show that there are not

more than two such sets. For if D_{13} is a third, D_{13} has at least one limit point on $[amb]$, we may suppose this to be the point m, and at least one limit point on K, since otherwise $[amb]$ would disconnect C. We choose on $[amb]$ four points in order $ap'q'mr's'b$. We construct, as in the foregoing paragraphs, two arcs phs and qkr, such that $[phs] \subset D_{11}$ and $[qkr] \subset D_{12}$, and we have on amb the order $apqmrsb$. Then, precisely as before, the simple closed curve $K'' = phsrkqp$, where sr and qp are the subarcs of amb, does not disconnect C. Therefore if K is any simple closed curve of C, and D_1 is either of its domains in C, and $[amb]$ is an arc such that $[amb] \subset D_1$ while a and b are on K separating two points o_1 and o_2, then $D_1 - [amb]$ is the sum of two subdomains† D_{11} and D_{12}, and the boundary‡ of D_{1i} is the simple closed curve $ambo_i a (i = 1, 2)$.§

6.2. Let x be any point of K and $G'' = \sum_{i=1}^{k} G_i$ a finite set of subcontinua of C which $\frac{1}{2}\epsilon$-separate x in C, $\epsilon < d(K)$.‖ There exist subcontinuous curves $F_i (i = 1, 2, \cdots, k)$ of C such that $F_i \supset G_i$, and $d(F_i) \leqq d(G_i) + \frac{1}{2}r''$, where $2r'' \leqq r(x, G'') \leqq \frac{1}{2}\epsilon$, and $2r'' \leqq r(G_i, G_j)$, $1 \leqq i < j \leqq k$.¶ It is clear that $F_i \times F_j = 0$, $i \neq j$. Since $G'' \subset U_{x(1/2)\epsilon}$, it follows that $F'' = \sum_{i=1}^{k} F_i \subset U_{x\epsilon}$. By our choice of r'', the complement in C of F'' contains x. By our choice of ϵ, it contains a point x' of K such that $r(x, x') > \epsilon$; x' does not belong to the Comp_x (rel. G'').†† Since $F'' \supset G''$, $C - F''$ is not connected. It is seen that F'', which is the sum of k disjoint continuous curves, ϵ-separates x.‡‡

Suppose $x' \subset \mathrm{Comp}_x(\text{rel. } F_i)$; $i = 1, 2, \cdots, k$. We discard from the sequence of sets $F_1, F_2, \cdots, F_k$ any curve F_i $(1 \leqq i \leqq k)$ such that F_i does not belong to $\mathrm{Comp}_x(\text{rel. } F_j)$, $j < i$. We reverse the order of the diminished sequence and repeat this process. We then have a sequence of sets, F_1', $F_2', \cdots, F_n'$ $(n \leqq k)$, such that $F_i' \subset \mathrm{Comp}_x(\text{rel. } F_j')$ if $i \neq j$. We add to each F_i' the set of all F_i'-domains excepting $\mathrm{Comp}_x(\text{rel. } F_i')$, $1 \leqq i \leqq n$. The resulting continua, $F_1^*, F_2^*, \cdots, F_n^*$, do not disconnect C. It is clear that $\sum_{i=1}^{n} F_i^* \supset \sum_{i=1}^{n} F_i'$. Also, $\sum_{i=1}^{n} F_i^* \supset \sum_{i=1}^{k} F_i$. For if $F_j (1 \leqq j \leqq k)$ is not contained in $\sum_{i=1}^{n} F_i^*$, $F_j \subset \mathrm{Comp}_x(\text{rel. } F_i')$, $i = 1, 2, \cdots, n$; and it is readily shown that F_j cannot have been discarded from the original sequence on

† Compare R. L. Moore, *On the foundations of plane analysis situs*, these Transactions, vol. 17 (1916), p. 144, Theorem 27.

‡ The boundary of a set X is the set $\overline{X} - X$.

§ In the argument of §6.1 we have made no use of the boundedness of C. In anticipation of the succeeding sections, see §9.

‖ See P. Urysohn, §3 loc. cit.

¶ By the method of §2, writing G_i for B.

†† Read the component relative to G'' which contains x.

‡‡ Compare with this section G. T. Whyburn and W. L. Ayres, *On continuous curves in n dimensions*, Bulletin of the American Mathematical Society, vol. 34 (1928), pp. 349–360, Theorems 1 and 2.

either the first or second traversing of that sequence. Since $x+x' \subset C-\sum_{i=1}^{n} F_i^*$, $C-\sum_{i=1}^{n} F_i^*$ is not connected. But $F_1' \times F_2' = 0$, and $F_2' \subset \text{Comp}_x(\text{rel. } F_1')$. Since $C-F_1^* = \text{Comp}_x(\text{rel. } F_1')$, $F_1^* F_2' = 0$. Then $F_1^* \subset \text{Comp}_x(\text{rel. } F_2')$ and $F_1^* \times F_2^* = 0$. The argument holds for any i and j, $1 \leqq i < j \leqq n$. We shall show that the sum of a finite number of disjoint continua, no one of which disconnects C, cannot disconnect C.

Let L' be any arc of C joining a point of F_1^* to a point of one of the other continua, and L the subarc from the last point f on F_1^* to the first point f' thereafter on any of the other continua, say F_2^*. Since L does not disconnect C, and $L \times F_1^* = f$, it is clear that the continuum $F_1^* + L$ does not disconnect C. Similarly, $F_2^* + L$ does not disconnect C. But the product of these continua is connected: $(F_1^* + L) \times (F_2^* + L) = L$. Then the continuum $F_1^* + L + F_2^*$ cannot disconnect C. But by this argument we have established the first case for an inductive proof, and reduced the number of the continua by one. Therefore there is a single continuous curve of $\sum_{i=1}^{n} F_i$ which separates x and x'.

6.3. Since x is a non-cut point of C, there is a finite set of arcs $\sum_{i=1}^{n} L_i$ in $C-x$, which join the continuous curves F_j $(j = 1, 2, \cdots, n)$ to each other and to x'. Then $G = F'' + \sum_{i=1}^{n} L_i + x'$ is a continuum, and $r(x, G) > 0$. Since x is an interior point of $\text{Comp}_x(\text{rel. } G)$, there is a $\overline{U}_{x\delta} \subset \text{Comp}_x(\text{rel. } G)$. Let F^0 be the sum of a finite number of continuous curves which δ-separate x. If y is any point such that $r(x, y) > \epsilon$, any arc xy of C has at least one point in common with $F'' \subset G$. Then xy has a first point g after y, which is a point of G. If yg has any point on F^0 it has a first point f'' on F^0 such that yf'' contains no point of G. Since $F^0 \subset \text{Comp}_x(\text{rel. } G)$, $y \subset \text{Comp}_x(\text{rel. } G) \subset \text{Comp}_x(\text{rel. } F'')$; but F'' ϵ-separates x. Then $yg \times F^0 = 0$. Then y belongs to the same component relative to F^0 as does G, and consequently as does x'.

There is, by the preceding section, a single continuous curve F of F^0 which separates x and x'. But if y is any point such that $r(x, y) > \epsilon$, $y \subset \text{Comp}_{x'}(\text{rel. } F^0) \subset \text{Comp}_{x'}(\text{rel. } F)$, and F separates x and y. Then F ϵ-separates x.

6.4. Then x is an avoidable point:† i.e. for any ϵ there is a δ_ϵ such that if $y+z \subset U_{x\delta_\epsilon}$, there is an arc yz of $C-x$, and $d(yz) < \epsilon$. If y and z are chosen to separate x and x' on K, the arc yz has a last point a on xyx' and a first point b thereafter on xzx', and the subarc ab has only the points a and b on K. Then $[ab]$ belongs to one of the two K-domains. We can construct an infinite sequence of such arcs converging to x. An infinite subsequence of the open arcs belong to the same one of the K-domains, say $\mathcal{D}_1$; let the sequence

† A "*vermeidbarer Punkt.*" See P. Urysohn, §3 loc. cit., Theorem 5.

be $a_1b_1, a_2b_2, \cdots, a_nb_n, \cdots$. The arc a_1b_1 divides D_1 into two subdomains such that one of these, D_{11}, has the boundary $a_1b_1xa_1$ $(K-b_1xa_1 \supset x')$, while x is not a limit point of the other domain. Since the sequence of arcs converge to x, and are contained in D_1, all but a finite number of them belong to D_{11}. Say a_2b_2 is the first which is contained in D_{11}. Then a_2b_2 divides D_{11} into two subdomains such that one of these, D_{12}, has the boundary $a_2b_2xa_2$ $(b_2xa_2 \subset b_1xa_1)$. We construct an infinite sequence of such domains, $D_1 \supset D_{11} \supset D_{12} \supset D_{13} \supset \cdots$; the boundary of D_{1n} is an arc a_nxb_n of K and an arc a_nb_n of D_1. Let y be a point of D_1-D_{11}. Then y is not contained in any D_{1n} $(n=1, 2, \cdots)$. Suppose there is an ϵ'' such that $d(D_{1n})>\epsilon''$; then $\prod_{n=1}^{\infty} D_{1n}$ defines a closed connected set of diameter at least ϵ''. If y' is any point of this set, every arc yy' of D_1 has at least one point in common with the boundary of every D_{1n}, and therefore at least one point in common with the arcs $a_nb_n(n=1, 2, \cdots)$ and consequently it must contain the point x since these arcs converge to x.† But x is not a point of D_1. Therefore no such ϵ'' exists, and for any preassigned ϵ there is an n_ϵ such that $d(D_{1m})<\epsilon$ when $m \geqq n_\epsilon$.

6.5. For a preassigned ϵ, $\epsilon<d(K)$, let D_{1n} be one of the domains, defined above, of diameter less than ϵ. There is a d_ϵ such that $U_{xd_\epsilon} \times D_1 \subset D_{1n}$. Let F be a continuous curve which d_ϵ-separates x in C. Then on each arc xx' of K (the x' of the preceding sections), from x', there is a first point h and k respectively which belongs to F. Then $hxk \supset K \times F$. Moreover, $F+hxk$ is a continuous curve. Also, $F-[hxk]$ is not connected, since there are points of F in D_2 and in $D_{1n} \subset D_1$. Let $J=hxk+F \times D_2$. Since $F \times D_2$ is an open subset of F it consists of a number of domains relative to hxk in the curve $F+hxk$; by the Remark to the Lemma, it follows that J is a continuous curve. The closed set $\overline{D}_{1n}$ does not disconnect C. Also $J \times \overline{D}_{1n} = hxk \times a_nxb_n$ and is connected. But $J+\overline{D}_{1n} \supset F$. Since x and x' do not belong to F, there are in D_2 a point $p \subset \mathrm{Comp}_x$(rel. F) and a point $q \subset \mathrm{Comp}_{x'}$(rel. F), and $(p+q) \times (J+\overline{D}_{1n})=0$. Then the complement of $J+\overline{D}_{1n}$ is not connected. Therefore J disconnects C. There are two points h' and k' of $K \times F$, in order $hh'xk'k$, such that no point of $[h'xk']$ belongs to F. We wish to show that $J-[h'xk']$ is connected. Otherwise it is the sum of two continuous curves J_1 and J_2. We form J_1^* and J_2^* as previously, adding to J_1 and to J_2 respectively those respective J_1-domains and J_2-domains which do not contain x, therefore not D_2 and not x'. It follows readily (compare §4.1 and §6.2) that $J_1^* \times J_2^*=0$; that $J_1^*+h'xk'+J_2^* \supset J$, and disconnects C; moreover, that this is impossible.

† It is understood that the set of arcs need not be the set originally defined, but a suitable subsequence determined by the preceding construction.

Therefore $J-[h'xk']$ is connected. Then it is a continuous curve.† There is an arc $h'k'$ of $J-[h'xk']$, and this cannot be a subset of K since it contains neither x nor x'. Then this has a last point a' on $xh'x'$ and a first point b' thereafter on $xk'x'$, and $[a'b'] \subset J-J\times K=F\times D_2 \subset D_2$. Then, precisely as in §6.4, for any preassigned ϵ we can find a domain D_{2n}, such that $D_{2n} \subset D_2$, $d(D_{2n})<\epsilon$, and $\overline{D}_{2n}-D_{2n}=a_n' b_n'+b_n' xa_n'$. Also, if D_2 contains any sequence of which x is the sequential limit point, all but a finite number of the points of this sequence belong to D_{2n}.‡

From the existence of the domains defined above, it is clear that $x+D_i$ $(i=1, 2)$ is connected im kleinen. Then x is accessible§ from D_i $(i=1, 2)$. For any preassigned ϵ'' there is a domain (see §6.4 and the preceding paragraph) D_{1n} of D_1 and a domain D_{2n} of D_2 such that $d(D_{in})<\frac{1}{2}\epsilon''$ $(i=1, 2)$, $K\times\overline{D}_{1n}=a_nxb_n$ and $K\times\overline{D}_{2n}=a_n' xb_n'$. Since every point of K (the arguments are not peculiar to any point of K) is accessible from each of the domains of K, and from any domain of which it is the boundary point (since any simple closed curve may replace K), we may suppose that the arcs a_nxb_n and $a_n' xb_n'$ coincide; i.e., $a_n'=a_n\equiv a$, and $b_n'=b_n\equiv b$. Let ao_1b be that arc of D_1 which is on the boundary of D_{1n}, and ao_2b the arc of D_2 which is on the boundary of D_{2n}. Then $K''=ao_1bo_2a$ divides C into two domains D_1'' and D_2''; let $D_1''\supset x$. It is readily seen that D_1'' contains $[axb]$ and the two domains D_{1n} and D_{2n}; moreover, that $D_1''=D_{1n}+[axb]+D_{2n}$. Then $d(D_1'')\leqq d(D_{1n})+d(D_{2n})<\epsilon''$. Therefore K'' ϵ''-separates x. Since every point z of C belongs to at least one simple closed curve of C, it follows that every point of C can be ϵ-separated, for any ϵ, by a simple closed curve. This is our first assurance that the dimension of C cannot exceed two.

6.6. Let o' be any point of C, and call $C-o'$ the space S. We shall show that S satisfies all the axioms of R. L. Moore's paper‖ and is homeomorphic with the euclidean plane.¶ Every simple closed curve R' of S belongs to C and divides C into two domains, one of which contains o', the other not. We call the second domain, free of o', the region R of S for the simple closed curve R'; every point of it belongs to S. The domain of R' in C which contains

† Compare Theorem 1 of Gehman's thesis, Annals of Mathematics, vol. 27 (1925), pp. 29–46.

‡ For the case that C is bounded, the method of this section, replacing D_{1n} by D_1, is applicable to either domain of K. The arguments are constructed to eliminate repetition in the treatment of the unbounded case; see the first sections of Theorem 6.

§ See G. T. Whyburn, *Concerning accessibility in the plane and regular accessibility in n dimensions*, Bulletin of the American Mathematical Society, vol. 34 (1928), p. 509. The regularity of accessibility is implicit in our method. Compare Whyburn's proof.

‖ See §6.1 loc. cit., p. 131.

¶ See R. L. Moore, *Concerning a set of postulates for plane analysis situs*, these Transactions, vol. 20 (1919), pp. 169–178.

o' is called the exterior of the region R in S. Every point x of S is a point x of C and can be ϵ-separated in C, for preassigned ϵ, by a simple closed curve of C. If $r(x, o') = \delta_x$, and R_x' $\frac{1}{2}\delta_x$-separates x in C, then R_x' does not contain o' and belongs to S, and the domain of R_x' in C which contains x cannot contain o'; therefore it is the region R_x of S. Then *every point of S belongs to a region of S which is of diameter less than a preassigned* ϵ.

6.7. Let $\{K^1\}$ be a set of regions of S of diameter less than a preassigned ϵ_1, such that every point x of S belongs to at least one region of the set, and such that if $r(x, o') = \delta_x$, then every region of the set which contains x is of diameter less than $\frac{1}{2}\delta_x$.† From the Lindelöf theorem it follows that there is a countable subset of $\{K^1\}$ which covers S. We modify this subsequence as follows: If K_{11} is the first region, we discard from the sequence every succeeding region which is contained in K_{11}. If K_{12} is the first remaining region, we discard among its successors (in the new sequence) all regions which are contained in $K_{11}+K_{12}$. Continuing indefinitely we arrive at a sequence $K_{11}, K_{12}, \cdots, K_{1n}, \cdots$; which cannot be vacuous, since it contains at least the one region K_{11}; which covers every point x of S, or it is readily seen that a region of the original set was discarded which should not have been; no region of it is contained in the sum of the preceding regions; no region covers o' nor is o' on the boundary of any region; finally, the sequence cannot be finite. We construct a countable sequence of such sequences: $K_{11}, K_{12}, \cdots, K_{21}, K_{22}, \cdots, K_{n1}, K_{n2}, \cdots$; where $d(K_{ni}) < (1/2)^{n-1}\epsilon_1$, $i = 1, 2, \cdots$. We order this countable set in traditional fashion: if K_{ij} and K_{pq} are two regions of this set, K_{ij} precedes K_{pq} if $i+j < p+q$ while if $i+j = p+q$ then K_{ij} precedes K_{pq} if $i < j$. We renumber this, $K_1, K_2, \cdots, K_n$, and call it our fundamental sequence of regions.

Consider now any point x of S. In the set of regions $\sum_{i=1}^{\infty} K_{1i}$ every region which contains x is of diameter less than $\frac{1}{2}\delta_x$ (defined as above). Let M be the set of points of S whose distance from x is not greater than $\frac{3}{4}\delta_x$. Then M does not include o', and is closed. Then in the sequence $\sum_{i=1}^{\infty} K_{1i}$, there is a finite set of regions which covers M. If among the second indices of this finite set the index m is the largest, then $M \subset \sum_{i=1}^{m} K_{1i}$. If now K_{1k} $(k > m)$ contains x, it is clear that $K_{1k} \subset M \subset \sum_{i=1}^{m} K_{1i}$. Since this contradicts our construction of the sequence, it is clear that at most a finite number of regions of the first sequence can contain x. The argument is valid for any sequence $\sum_{i=1}^{\infty} K_{ni}$ $(n = 1, 2, \cdots)$. Then in the fundamental sequence there are at most a finite number of regions which contain a given point x and are of diameter greater than a preassigned ϵ. From this, and the con-

† This is to be understood as true for *every* point in any region of the set.

struction of our fundamental sequence, it is readily shown that this sequence satisfies, for S, Axiom 1 of Moore's paper.

The regions of S are connected point sets; this is Axiom 2. If K is any simple closed curve of C and x a point of one of the K-domains, D_1 say, then $K+D_2$ (the other domain) does not disconnect C; therefore $K+D_2+x$† cannot disconnect C, and D_1-x must be connected. But the exterior of a region S is some domain of C minus the point o'. Then the exterior of a region of S is connected (Axiom 3).‡ A region and its boundary is a closed and bounded subset of a continuous curve; therefore it has the Heine-Borel property (Axiom 4). Any sequence of points of S such that the same sequence on C has the sequential limit point o', has on S no limit point (Axiom 5). The Axioms 6 and 7 are readily derived from the existence of the domains defined in §6.4 and §6.5. By our definition, every simple closed curve of S determines a region of S (Axiom 8).

6.8. Then S is homeomorphic with the euclidean plane.§ But this is homeomorphic with the complement on the surface of a sphere S' of a single point o of S'. Then $C-o'$ is homeomorphic with $S'-o$, and the homeomorphism must extend to C and to S', if we merely add to it that it transforms o' into o and reciprocally. Therefore C is a simple closed surface, and the theorem is proved.

7. If T is an acyclic continuous curve, bounded or unbounded, there is in T a unique arc xy for any two given points x and y. Then if I_1 is a subcontinuum of T which separates x and y it must contain at least one point of this arc. Therefore if I_1 and I_2 are two continua of T neither of which separates x and y, their sum cannot separate x and y. It follows that if I_1 and I_2 are two continua of T neither of which disconnects T, their sum cannot disconnect T. On the other hand, the product of two continua of T is always connected (or vacuous). Therefore T may be said to satisfy the Janiszewski-Mullikin Theorem. But since T never contains two continua whose product fails to be connected, we shall say that it satisfies this theorem *vacuously*. It is clear that no continuous curve which contains at least one simple closed curve can satisfy this theorem vacuously, in the above sense.

8. We prove the following theorem.

---

† See Remark to Theorem 2. A similar argument obtains.

‡ In a later connection it will be observed that this argument is valid, even if C is unbounded, provided D_2 is bounded. Moreover, this is the only case that need concern us, since regions of S are bounded. See the first sections of Theorem 6.

§ See §6.6 second note.

THEOREM 5. *C is a continuous curve satisfying non-vacuously the Janiszewski-Mullikin Theorem. Then the maximal cyclicly connected continuous curves of C are simple closed surfaces.*†

Since C cannot be an acyclic continuous curve, let K be any simple closed subcurve and J the maximal cyclicly connected continuous curve of C which contains K. If J is C our theorem is proved. We shall show that J satisfies the Janiszewski-Mullikin Theorem. There are two cases to consider.

(1) Suppose that I_1 and I_2 are two subcontinua of J neither of which disconnects J and such that I_1+I_2 disconnects J although $I_1\times I_2$ is connected. We form I_1^*, as previously, adding to I_1 all components in C relative to it, excepting that one which contains $J-I_1$; also, I_2^* by adding to I_2 all components of $C-I_2$ excepting that one which contains $J-I_2$. Then if p and q are points of $J-(I_1+I_2)$ not in the same component in J relative to (I_1+I_2), neither p nor q belongs to $(I_1^*+I_2^*)$, and therefore $C-(I_1^*+I_2^*)$ is not connected. Since $I_1^*\times I_2^*\supset I_1\times I_2$, and $I_1\times I_2$ is connected, to show that $I_1^*\times I_2^*$ is connected it will be sufficient to show that if m is any point of $I_1^*\times I_2^*$ there is a connected set H, such that $m\subset H\subset I_1^*\times I_2^*$ and H has at least one limit point on $I_1\times I_2$. If $m\subset I_1\times I_2$, let $H=m$. Suppose then that m is not a point of I_1. Since $m\subset I_1^*$, while $J-I_1\subset C-I_1^*$, m is not a point of $J-I_1$ and therefore not a point of J. Then m belongs to a component H relative to J in C. Then H has a single limit point m' on J. If $m'\subset J-I_1$, $(J-I_1)+H$ belongs to one component of $C-I_1$, and H and therefore m cannot belong to I_1^*, since this component is not added to I_1. Therefore $m'\subset I_1$. Similarly $m'\subset I_2$. Then the component of $C-I_1$ containing $J-I_1$ cannot also contain H since every connected set which has a point on $J-I_1$ and a point on H must contain m'(or we find in the complement of J a connected set which has two limit points on J; this is not possible).‡ Then H must have been added to I_1 and $H\subset I_1^*$. Similarly, $H\subset I_2^*$, and $H\subset I_1^*\times I_2^*$. Then we have found in C two continua, I_1^* and I_2^*, neither of which disconnects C, whose product is connected and whose sum disconnects C. But C satisfies the Janiszewski-Mullikin Theorem.

(2) Suppose that I_1 and I_2 are subcontinua of J such that $J-I_1$ and $J-I_2$ are connected, that $I_1\times I_2$ is not connected, but that $J-(I_1+I_2)$ is connected. We form I_1^* and I_2^* as above. If $I_1^*\times I_2^*$ is connected, it contains an irreducible continuum G which contains p and q of $I_1\times I_2$, and p and q belong to

† A maximal cyclicly connected continuous curve J of C is a cyclicly connected continuous subcurve J of C and is unique for a given simple closed curve K of C. The components in C relative to J have a single limit point on J. See G. T. Whyburn, §1 loc. cit.

‡ See G. T. Whyburn, above, Theorem 2.

no connected subset of $I_1 \times I_2$. Then G cannot belong to $I_1 \times I_2$. Say m of G is not a point of $I_1 \times I_2$. As in the first case, $m \subset H \subset I_1^* \times I_2^*$, and H is a component of $C-J$. Then $G-(G \times H)$ is not vacuous since it contains $p+q$, is a proper subset of G since it fails to contain m, and is closed because $G \times H$ is an open subset of G. If $G-(G \times H) = M+N$, $M \times \overline{N} = \overline{M} \times N = 0$, let M contain m', where m' is the unique limit point which H has on J; this is a point of G, since every continuum joining m and p must contain m'. Then $G = (M+G \times H)+N$ and is not connected. Therefore $G-(G \times H)$ is a proper subcontinuum of G joining p and q, although G was assumed irreducible between p and q.

If now m is a point of $C-(I_1^*+I_2^*)$ it belongs to $C-(I_1+I_2)$, and therefore either to $J-(I_1+I_2)$ or to a component H of $C-J$ which has a single limit point m' on $J-(I_1+I_2)$; and therefore the complement of $(I_1^*+I_2^*)$ is connected.†

9. We prove the following theorem:

THEOREM 6. *C is a cyclicly connected unbounded continuous curve satisfying the Janiszewski-Mullikin Theorem. Then C is homeomorphic with the complement on a simple closed surface of a closed and totally disconnected point set.*

It is not possible, given an arbitrary continuum X such that $C-X$ is not connected, to form the continuum X^* which does not disconnect C and to argue a contradiction on X^* as we did in the preceding theorems. For it may happen that X^* is unbounded, and our hypothesis is not applicable. We can anticipate our use of this process so that it is available when we have need of it.

If ab is an arc of C such that $C-ab$ is not connected and if q is any point of $[ab]$ either aq or qb must disconnect C. If ab is expressed as the sum of any finite number of arcs, two of which have at most an end point in common, at least one of these arcs disconnects C. But for any preassigned ϵ, ab can be expressed as the sum of a finite number of arcs no one of which is of diameter greater than ϵ. We can construct a sequence (t) of such arcs, each a subset of the preceding, which converge to a point z of ab (z may be a or b). Since the number of unbounded ab-domains is finite, there is a finite set L'' of arcs of $C-z$ which join all of these domains. There is an arc of the sequence (t), call it xy, such that $xy \times L'' = 0$. Relative to xy there is a single

† Owing to the length of this paper it has seemed advisable to omit details which are essentially repetitions of previous argument, or too readily supplied by one familiar with the field of analysis situs.

unbounded domain; call this N_1. For this arc xy, the argument of §6 is valid for C unbounded. Then §6.1 is also valid (see §6.1 third note).

If, now, x is any point of a simple closed curve K of C, we ϵ-separate x in C by the sum of a finite number of continuous curves, precisely as in the first paragraph of §6.2. By the argument of §6.3, omitting the last paragraph, we find for x a corresponding ϵ-separating set relative to which there is a single unbounded domain, and this contains the point x' of that argument. For this ϵ-separating set the arguments of §§6.2, 6.3, 6.4, and 6.5 are valid independently of the boundedness or unboundedness of C. We define a region of C, for the simple closed curve R' of C, as the bounded domain *if it exists* of R'. Replacing S by C, and omitting all reference to the point o', the arguments of §§6.6 and 6.7 are valid to show that C satisfies all the axioms of Moore's paper except the eighth. In general, there will be simple closed curves of C both of whose domains on C are unbounded; these determine no region in the sense of our definition. However, if R is a region of C and K' is any simple closed curve of R, K' has a bounded domain and determines a region which is a subset of R. It is seen that R satisfies all the axioms of Moore's paper, and is homeomorphic with the plane. When, therefore, our arguments are confined to a region of C we are at liberty to avail ourselves of any plane theorem.

9.1. We shall show that if o' is a point and F a continuum (bounded) which does not contain o', there is a finite set of simple closed curves of C which separate o' and F, and whose upper distance from F does not exceed a preassigned ϵ. Cover F by regions $\{K\}$ of diameter less than ϵ, such that none of these contains o' or has o' on its boundary. There is a finite covering set $K_1, K_2, \cdots, K_m$, whose boundaries we denote by $K_1', K_2', \cdots, K_m'$. Let $F''=\sum_{i=1}^{m} K_i$, and $K''=\sum_{i=1}^{m} K_i'$; then $F'' \supset F$, and $K'' \supset \overline{F}''-F''$. Let $H=\mathrm{Comp}_{o'}(\text{rel. } F'')$, and let x be any point of $\overline{H}\times K''$. Let R_x be a region of C containing x, such that R_x does not contain o' nor all points of any of the set K'' of simple closed curves, and let R be a subregion of R_x containing x, such that $\overline{R} \subset R_x$. We shall regard R_x as a euclidean plane to the extent that we are free to use within it plane theorems; when convenient, we shall think of it as a region of C.

Consider $J=R'+R\times K''$; it is a continuous curve.† Since every point k of $R\times K''$ belongs to a simple closed curve of K'' not every point of which is contained in $\overline{R}$, k is interior to an arc hkm of K'' such that h and m are distinct points on R', and $[hkm] \subset R\times K''$. It readily follows that J is cy-

† See the argument on the J of §6.5; R' is the boundary of R and is a simple closed curve, from our definition of region.

clicly connected. Let $r(x, R') = \rho_x$; x is a sequential limit for points $x_1, x_2, x_3, \cdots, x_n, \cdots$, of $H \times R \times U_{x(1/2)\rho_x}$. Each point x_n can be joined to o' by an arc of H, and this has a subarc $x_n x_n'$ such that $x_n x_n' - x_n' \subset R$, while $x_n' \subset R'$. Then $d(x_n x_n') \geqq \frac{1}{2}\rho_x$. It is seen that every point of $H \times R \times U_{x(1/2)\rho_x}$ belongs to a complementary domain of J in R_x consisting of points of $R \times H$, and of diameter at least $\frac{1}{2}\rho_x$. By a theorem due to Schoenflies, at most a finite number of these can be distinct. Then x is on the boundary of at least one complementary domain D_x of J (in R_x), and $D_x \subset H \times R$; the boundary of D_x is a simple closed curve.† This simple closed curve being a subset of J contains no point of H which is not a point of R'. On the other hand, it must contain at least one point of $H \times R'$, since $H \supset D_x + o'$ and is connected. Then there is an arc $[pxq] \subset \overline{H} \times R \times K''$, and $p+q \subset R'$, this arc being a subset of the boundary of D_x. If x is on the boundary of another complementary domain B_x of J consisting of points of H and R, choose the points d in D_x and b in B_x. There are arcs dx and bx such that $dx - x \subset D_x$, and $bx - x \subset B_x$. There is an arc db of H, which may be supposed to have only the points d and b in common with $dx + xb$.‡ Then $Q' = db + bx + xd$ is a simple closed curve of $H + x$. Since every arc of R joining a point of $dx - x$ to a point of $bx - x$ must contain at least one point of $J - R' \subset K''$, it follows from the existence of the arcs defined in §§6.3 and 6.4 that there are points of K'' in each of the Q'-domains on C. Since x is the only point of K'' on Q', it follows that there is at least one simple closed curve of K'' in each Q'-domain (excepting perhaps the point x); therefore at least one region of F'', and therefore at least one point of F in each Q'-domain. But Q' contains no point of F, and F is connected. This is clearly not possible. If, now, (x_i) is a set of points of $\overline{H} \times K'' \times R \times U_{x(1/2)\rho_x}$ of which x is the sequential limit point, each of these is on the boundary of a complementary domain of J consisting of points of $H \times R$. But in a sufficiently small neighborhood of x there are points of only one such domain, D_x. Then all but a finite number of the points of (x_i) belong to the boundary of D_x, and in consequence to pxq.

Then it is seen that x is of order two on $\overline{H} \times K''$. But x belongs to a maximal connected subset, necessarily closed, of $\overline{H} \times K''$; call this K_x'. It is apparent that every subcontinuum of K'' is a continuous curve. Therefore K_x' is a continuous curve. Since every point of it is of order two, K_x' is a simple closed curve. Then $\overline{H} \times K''$ is a set of simple closed curves. If the number of these is not finite, let (x') be a set of points of $\overline{H} \times K''$ not more

† See G. T. Whyburn, §1 loc. cit., p. 37, Theorem 10.

‡ This is not essential, although convenient. Compare R. L. Moore, §6.1 first note, p. 147, Theorem 32.

than a finite number of which belong to any single simple closed subcurve. Then if x' is any limit point of this set, $x' \subset \overline{H} \times K''$, and therefore to a simple closed curve $K_{x'} \subset \overline{H} \times K''$. But in a sufficiently small neighborhood of x' every point of $\overline{H} \times K''$ belongs to an arc $p'x'q'$ of $K_{x'}$, and therefore infinitely many of the points of (x') belong to this arc. This contradicts our choice of (x'). Therefore $\overline{H} \times K''$ is a finite set of simple closed curves. It is readily seen that $\overline{H} \times K''$ separates o' and F.

9.2. We form the inverse C^* of C with respect to a center of inversion v which is a point of the embedding euclidean space but not of C. In C^* we choose a $\frac{1}{2}\epsilon$-separating set $F^* = \sum_{i=1}^{n} F_i^*$ for v, such that $2\epsilon < d(C^*)$. Let o^* be a point of C^* such that $r(o^*, v) > \epsilon$. Then F^* does not contain o^* and separates o^* and v. Consider on C the corresponding point o, and the corresponding finite set of continua $F = \sum_{i=1}^{n} F_i$. By the preceding section there is a finite set of simple closed curves which separate o and F_1, and are at an upper distance from F_1 not greater than δ', where δ' is chosen (as is possible since inversion is a (1, 1) reciprocal bicontinuous transformation of C into $C^* - v$) so that the corresponding simple closed curves on C^* are at an upper distance from F_1^* less than $\frac{1}{2}\epsilon$. It is clear that these curves are contained in $U_{v\epsilon}^*$. Inductively there is a finite set of simple closed curves $K_{11}, K_{12}, \cdots, K_{1n}$, which separate o and F, such that the corresponding set $K_1^* = \sum_{i=1}^{n_1} K_{1i}^*$ is contained in $U_{v\epsilon}^*$. Since every arc o^*v contains at least one point of F^*, it has (from o^*) a first point f^* on F^*. The corresponding arc fo on C contains at least one point of $K_1 = \sum_{i=1}^{n_1} K_{1i}$, and therefore f^*o^* contains at least one point of K_1^*. Therefore K_1^* separates o^* and v. Let $M_1^* = \mathrm{Comp}_v(\text{rel. } K_1^*)$. We define a sequence of sets, $K_1^*, K_2^*, K_3^*, \cdots, K_m^*, \cdots$, of simple closed curves $(K_m^* = \sum_{i=1}^{n_m} K_{mi}^*)$ such that the sequence converges to v, each set separates o^* and v, and $M_n^* = \mathrm{Comp}_v(\text{rel. } K_n^*) \supset M_{n+1}^* + K_{n+1}^*$. If $\prod_{n=1}^{\infty} M_n^* \supset v'$ distinct from v, every arc o^*v' contains v; but v is not a cut point of C^*. Then $\prod_{n=1}^{\infty} M_n^* = v$.

9.3. We may suppose that no $n_1 - 1$ of the simple closed curves of K_1^* separate o^* and v (similarly for K_m^*). Consider, now, on C a region R_0 containing o, such that $\overline{R}_0 = R_0 + K_0$ (a simple closed curve) has no point in common with $K_1 = \sum_{i=1}^{n_1} K_{1i}$. Then K_1 lies in E, the exterior of R_0. For each curve K_{1i} $(i = 1, 2, \cdots, n_1)$ we call its "region" R_{1i} that one of its domains which does not contain o. It is readily seen that $\sum_{i=1}^{n_1} R_{1i} \subset E$. Let $E - \sum_{i=1}^{n_1} \overline{R}_{1i} = E - \sum_{i=1}^{n_1} (R_{1i} + K_{1i}) = D_0$. Suppose $D_0 = M + N$, and $M \times \overline{N} = \overline{M} \times N = 0$. Every point of K_0 is a limit point of E, but not of $\sum_{i=1}^{n_1} \overline{R}_{1i}$. Then every point of K_0 is a limit point of D_0. Suppose x of K_0 is a limit point of M. There is a subdomain D_x of E which has no point in common with $\sum_{i=1}^{n_1} \overline{R}_{1i}$, and whose

boundary has in common with K_0 an arc axb.† Then every point of D_x belongs to D_0, and since D_x is connected and x a limit point of it, $D_x \subset M$. Then every point of $[axb]$ is a limit point of M and not of N. From the connectedness of K_0 this is seen to be true for every point of K_0. If K_{1n_1} has any point in R_{11},‡ since $K_{1n_1} \times K_{11} = 0$, it is seen to be a subset of R_{11} and K_{11} separates o and K_{1n_1}. Then every arc ok_{n_1}, where k_{n_1} is any point of K_{1n_1}, has at least one point on K_{11} and the corresponding arc $o^*k_{n_1}^*$ on C^* has at least one point on K_{11}^*; then $\sum_{i=1}^{n_1-1} K_{1i}^*$ separates o^* and v. Then it follows that $K_{1n_1} \times R_{11} = 0$. It is seen that $K_{1n_1} \times (\sum_{i=1}^{n_1-1} \overline{R}_{1i}) = 0$. Then, precisely as for K_0, every point of K_{1n_1} is a limit point of M and not of N or vice versa. Similarly for any K_{1i} $(1 \leqq i \leqq n_1)$. We may suppose that every point of $\sum_{i=1}^{m} K_{1i}$ $(m \leqq n_1)$ is a limit point of M and not N, while every point of $\sum_{i=m+1}^{n_1} K_{1i}$ (this is vacuous if $m = n_1$) is a limit point of N and not M. Then

$$C = \left[\overline{R}_0 + \sum_{i=1}^{m} \overline{R}_{1i} + M\right] + \left[\sum_{i=m+1}^{n_1} R_{1i} + N\right]$$

and is not connected. Therefore D_0 is connected. It is seen that $C_1 = R_0 + D_0$ is connected.§

Let m^* be any point of $\text{Comp}_{o^*}(\text{rel. } K_1^*)$, on C^*. If the corresponding point m on C belongs to $\sum_{i=2}^{n_1} R_{1i}$, every arc mo has at least one point on K_1 and the corresponding arcs on C^* have at least one point on K_1^*. Therefore $m \subset C_1$. It is seen that C_1^*, corresponding on C^* to C_1, is the $\text{Comp}_{o^*}(\text{rel. } K_1^*)$. Since $\sum_{i=2}^{n_1} K_{1i}^*$ does not separate o^* and v, there is an arc o^*v in $C^* - \sum_{i=2}^{n_1} K_{1i}^*$ and this arc has at least one point in common with K_{11}^*. Then there is a point k_1^* on o^*v such that $(vk_1^* - k_1^*) \times K_1^* = 0$. The arc vk_1^* corresponds on C to a ray with the single point k_1 of K_{11} on K_1. Then this ray has no point in common with $\sum_{i=2}^{n_1} R_{1i}$ and cannot belong to D_0 which is bounded;‖ therefore it belongs to R_{11} which is therefore unbounded. Then the corresponding set R_{11}^* on C^* contains v, and $R_{11}^* \subset M_1^* = \text{Comp}_v(\text{rel. } K_1^*)$. Similarly $M_1^* \supset \sum_{i=2}^{n_1} R_{1i}^*$, and it is seen that $C^* = C_1^* + K_1^* + M_1^*$. Inductively, if $C_j^* = \text{Comp}_{o^*}$ (rel. K_j^*), $C^* = C_j^* + K_j^* + M_j^*$. Since $\prod_{n=1}^{\infty} (K_n^* + M_n^*) = v$, it follows that $C^* = \sum_{n=1}^{\infty} C_n^*$. Then $C = \sum_{n=1}^{\infty} C_n$, where C_n is the set on C corresponding to C_n^*, and is the $\text{Comp}_o(\text{rel. } K_n)$.

9.4. Returning to C, we may suppose, purely for its convenience, that n_1 of the preceding sections is equal to two: then $C = C_1 + K_1 + \sum_{i=1}^{2} R_{1i}$, where

† Compare the domains constructed in §§6.3 and 6.4.

‡ The argument is general, the subscripts merely convenient.

§ Combining this result with that of §9.1 it is readily found that $H \times K''$ of that section is a single simple closed curve.

‖ It is seen that $D_0^* \subset C_1^*$.

$K_1=\sum_{i=1}^{2} K_{1i}$, and $C_1=R_0+D_0+K_0$. We construct the arc k_1k_{11}: $[k_1k_{11}] \subset D_0$, $k_1 \subset K_0$, $k_{11} \subset K_{11}$. If, now, $D_0-k_1k_{11}=M+N$, $M\times\overline{N}=\overline{M}\times N=0$, by an argument parallel to that of the preceding section, it follows that $C-k_1k_{11}=M''+N''$, where $M'' \supset M$, $N'' \supset N$, and either $(K_0-k_1) \subset \overline{M}$ and $(K_0-k_1)\times\overline{N}=0$, or vice versa; similar relations obtain for K_1-k_{11}, and for K_2. Since no arc disconnects C, it follows that $D_0-k_1k_{11}$ is connected. We construct an arc k_2k_{12}: $[k_2k_{12}] \subset D_0-k_1k_{11}$, $k_2 \subset K_0$, $k_{12} \subset K_{12}$. As above, we deduce that $D_0-\sum_{i=1}^{2} k_ik_{1i}$ is connected. Let b_1 and b_2 be points of K_0 separating k_1 and k_2 on K_0, and b_1b_2 an arc of $D_0-\sum_{i=1}^{2} k_ik_{1i}$; let k'' be any point of b_1b_2. The simple closed curve $b_1k''b_2k_1b_1$ separates $K_{11}+(k_{11}k_1-k_1)$ from $K_{12}+k_{12}k_2$ and from o. Then its "region" R^{11} (in the sense of that domain which does not contain o) contains K_{11}. Similarly the "region" R^{12} of $b_1k''b_2k_2b_1$ contains K_{12}. It is seen that $R^{11} \supset R_{11}$, and that $R^{12} \supset R_{12}$. Then $R^{1i}-(\overline{R}_{1i}+[k_ik_{1i}])$ is connected, by the first part of §9.3, and contains an arc b_1b_{1i} where b_{1i} is a point of K_{1i} distinct from k_{1i} $(i=1, 2)$.

Consider a sphere S in euclidean three-space. Let J_0 be any great circle on S, and Q_0 a hemisphere of J_0 on S. On the other hemisphere of S choose two circles J_{11} and J_{12}, calling Q_{11} and Q_{12} those respective domains which do not contain Q_0. We complete on S the configuration above, writing J for K, j for k, Q for R, a for b, and S for C, and preserving the subscripts. If, on C, O_R is any point in the exterior of R_0, there is a sequence of simple closed curves, each† of which separates O_R and R_0, which converge to K_0. It is readily found that the domains of these simple closed curves converge to R_0, and that there is a first, therefore, which is bounded. Then R_0 is interior to some region of C and, this being homeomorphic with the plane, R_0 is homeomorphic with the interior and boundary of a plane circle. Then there is a homeomorphism T_0: $T_0(R_0)=Q_0$, and $T_0(K_0)=J_0$.

On K_{11} choose two points r_1 and r_2 separating k_{11} and b_{11} (on S replace r by q), and let r_1 be the point such that the subdomain R_{111} of R^{11} corresponding to the simple closed curve $b_1k_1k_{11}r_1b_{11}b_1$ does not contain R_{11}. By the method of Moore's paper,‡ there is a homeomorphism T_{11} which transforms R_{111} into Q_{111} (the corresponding domain on S), and preserves the correspondence T_0 on the arcs b_1k_1 of K_0 and a_1j_1 of J_0. Similarly, there is a homeomorphism T_{12} which transforms the subdomain R_{112} determined by the simple closed curve $b_1b_{11}r_2k_{11}k_1b_2b_1$ into Q_{112} (on S), and determines a correspondence of their boundaries which reduces to T_0 along k_1b_1 of K_0 and to T_{11} on the arcs k_1k_{11} and b_1b_{11} of R'_{111}. Inductively, we define a transformation $T_1=\sum_{i=1}^{4} T_{1i}$ such

† See §9.3 third note.

‡ See §6.6, second note.

that $(T_0+T_1)(C_1)=S_1$, $T_0(K_0)=T_1(K_0)=J_0$, and $T_1(K_{1i})=J_{1i}$, $i=1, 2$.

9.5. From the relation on C^*, $\sum_{i=1}^{2} R_{1i}^* = M_1^* \supset M_2^* + K_2^*$, it follows that on C the $\overline{R}_{2i}$ $(i=1, 2, \cdots, n_2)$ fall into groups each contained within a single R_{1i} $(i=1, 2)$. Then for K_{11} and those curves of $K_2=\sum_{i=1}^{n_1} K_{2i}$ which belong to R_{11} the construction of the corresponding homeomorphism, for the set of simple closed curves in Q_{11} on S, does not differ from the method we have set forth. We construct on S a sequence of sets of simple closed curves J_0, $J_1, J_2, \cdots$, corresponding to the sequence $K_0, K_1, K_2, \cdots$ on C, each set on S having that relation to the "regions" on S determined by the previous set, which obtains for the sequence on C. We choose the curves on S of diameters converging to zero, so that there is defined on S a closed and totally disconnected point set B. We define $S_n = \text{Comp}_{o'}(\text{rel. } J_n)$, where o' is a point of Q_0; it is seen that $S-B=\sum_{n=1}^{\infty} S_n$. Inductively, as above, we construct a sequence of homeomorphisms, $T_0, T_1, \cdots, T_n, T_{n+1}, \cdots$, such that $(\sum_{i=0}^{n} T_i)(\overline{C}_n)=\overline{S}_n$, and $T_n(K_n)=T_{n+1}(K_n)=J_n$. It is seen that $T=\sum_{i=0}^{\infty} T_i$ is a homeomorphism of C into $S-B$. Therefore C is homeomorphic with the complement on the surface of a sphere of a closed and totally disconnected point set. If B is a single point, it is apparent that C is a plane; if B consists of two points, C may be recognised as a cylinder unbounded at both ends. It can be seen that C is essentially a generalization of a cylindrical surface; it resembles a tree, moreover, in its effect of branching. For this reason the name cylinder-tree has seemed appropriate.

9.6. The analogy with the tree (acyclic continuous curve) can be made more precise. For, by the method of Moore's paper, there is constructed a set of rulings of $\overline{R}^{11}-R_{11}$ by simple closed curves, such that the sum of these curves is the set $\overline{R}^{11}-R_{11}$, and each curve has in common with b_1b_{11} a single point; and the set of these simple closed curves is upper semicontinuous, so that the arc b_1b_{11} is equivalent to $\overline{R}^{11}-R_{11}$ in the sense of a Zerlegungsraum. Correspondingly for the arc b_1b_{12} of $\overline{R}^{12}-R_{12}$. Also there is an arc ob_1 of $\overline{R}_0$, and an upper semicontinuous set of simple closed curves ruling $\overline{R}_0$, each having a single point on ob_1. Then it follows that the tree $ob_1+b_1b_{11}+b_1b_{12}$ is equivalent to C_1 in the sense of a Zerlegungsraum, and this construction can be continued inductively, to define on S an acyclic continuous curve with end points $B+o$, almost all of whose cut points correspond to simple closed curves of $S-B$, and therefore to simple closed curves of C. Such points as b_1 may correspond to a sum of three, in general any finite number greater than two, of arcs distinct except for their end points. This tree corresponds, by the homeomorphism, to an unbounded acyclic continuous curve on C with the single end point o, and it is equivalent to C as a Zerlegungsraum.

9.7. Suppose, now, that S is any sphere (surface) and B a closed and to-

tally disconnected subset. If I is a continuum of $S-B$ which does not disconnect S, then $I+B$ is an upper semicontinuous collection of continua no one of which disconnects S, and therefore their sum does not disconnect S.† It is seen that $(S-B)-I$ is connected. If I is a continuum of $S-B$ which disconnects S, there are two points p and q which belong to different complementary domains of I in S, and are not points of B (since B is nowhere dense on any domain of S). Then I disconnects $(S-B)$. If, now, I_1 and I_2 are two continua of $S-B$ neither of which disconnects $S-B$ and their product is connected, their sum is a continuum which cannot disconnect S, and therefore not $S-B$. If their product is not connected, their sum disconnects S, and therefore $S-B$. Then $S-B$ satisfies the Janiszewski-Mullikin Theorem, and our theorem has given a necessary as well as sufficient condition.

10. The corresponding theorems on an unbounded continuous curve C which is not cyclicly connected, and satisfies non-vacuously the Janiszewski-Mullikin Theorem, are exceedingly complicated by the possibility of existence on a given maximal cyclicly connected subcontinuous curve J of C of a set of points such that the complement in C of one of these points contains at least one unbounded component distinct from J. For this reason no proof as in Theorem 5 is possible. Using the method of Theorem 5 on other points of J, it has seemed necessary to establish the arguments of Theorem 6 on them, and then to argue exceptionally on these "singular" points. These are found to be a closed and isolated set, and it can be shown that J is a cylinder-tree. In default, however, of a sufficiently direct proof, and owing to the length of this paper, this case is not discussed.

† See R. L. Moore, *Concerning upper semicontinuous collections of continua*, these Transactions, vol. 27 (1925), pp. 416–428.

UNIVERSITY OF PENNSYLVANIA,
PHILADELPHIA, PA.

ON EXTENDED STIELTJES SERIES*

BY

HUBERT STANLEY WALL

1. Let

$$c_0 - c_1 z + c_2 z^2 - \cdots \tag{1}$$

be a power series with real coefficients such that the determinants

$$A_n = \begin{vmatrix} c_0, & c_1, & \cdots, & c_{n-1} \\ c_1, & c_2, & \cdots, & c_n \\ \cdots & \cdots & \cdots & \cdots \\ c_{n-1}, & c_n, & \cdots, & c_{2n-2} \end{vmatrix}, \quad B_n = \begin{vmatrix} c_1, & c_2, & \cdots, & c_n \\ c_2, & c_3, & \cdots, & c_{n+1} \\ \cdots & \cdots & \cdots & \cdots \\ c_n, & c_{n+1}, & \cdots, & c_{2n-1} \end{vmatrix},$$

$n=1, 2, 3, \cdots$, are all positive. Then we define a *kth extension* of (1) to be a series

$$(-1)^k \frac{c_{-k}}{z^k} + (-1)^{k-1} \frac{c_{-k+1}}{z^{k-1}} + \cdots - \frac{c_{-1}}{z} + c_0 - c_1 z + c_2 z^2 - \cdots \tag{2}$$

such that all the determinants formed from the A_n and B_n by replacing throughout c_i by c_{i-k}, $i=0, 1, 2, 3, \cdots$, are positive.

In a previous paper† in these Transactions the present writer gave a necessary and sufficient condition for the existence of a first extension of (1), and gave examples to show that for any k there are series possessing a kth but not a $(k+1)$st extension, and others possessing extensions of infinite order. The condition there given is as follows. Let

$$\frac{1}{a_1} + \frac{z}{a_2} + \frac{z}{a_3} + \cdots \tag{3}$$

be the Stieltjes‡ continued fraction corresponding to the Stieltjes series (1). Then if $\sum a_{2i} = a_2 + a_4 + \cdots$ converges, and only then, a first extension exists and we may choose $c_{-1} \geqq \sum a_{2i}$ at pleasure. If c_{-p} exists then c_{-p-1} exists if and only if the series§ $\sum a_{2i}^{-p}$ in the continued fraction

* Presented to the Society, December 31, 1928; received by the editors in February and April, 1929.

† H. S. Wall, *On the Padé approximants associated with the continued fraction and series of Stieltjes*, these Transactions, vol. 31 (1929), pp. 91–116, Chapter III.

‡ Stieltjes, *Recherches sur les fractions continues*, Annales de Toulouse, vol. 8, J, pp. 1–122, and vol. 9, A, pp. 1–47, 1894–95; or Oeuvres, vol. 2.

§ Here and hereafter I write the superscripts without parentheses.

$$\frac{1}{a_1^{-p}} + \frac{z}{a_2^{-p}} + \frac{z}{a_3^{-p}} + \cdots \tag{4}$$

corresponding to the Stieltjes series

$$c_{-p} - c_{-p+1}z + c_{-p+2}z^2 - \cdots \tag{5}$$

converges. The minimum value of c_{-p-1} is $\sum a_{2i}^{-p}$, $p=0, 1, 2, \cdots, a_n^0 \equiv a_n$.

It will be convenient to make the following definition. The kth extension of (1) in which every c_{-p}, $p=1, 2, 3, \cdots, k$, has its minimum value is the *minimal* kth extension of (1).

In the following article I shall give a necessary and sufficient condition for a minimal kth extension of (1), and then show that throughout a large class of Stieltjes series, including among others all those for which $\sum a_i = a_1 + a_2 + a_3 + \cdots$ converges,* minimal extensions of infinite order exist. Furthermore, if in this case we form the Stieltjes series

$$\frac{c_{-1}}{z} - \frac{c_{-2}}{z^2} + \frac{c_{-3}}{z^3} - \cdots \tag{6}$$

with corresponding Stieltjes continued fraction

$$\frac{1}{\alpha_1 z +} \frac{1}{\alpha_2 +} \frac{1}{\alpha_3 z +} \cdots \tag{7}$$

then the latter converges over any finite region not containing a part of the negative half of the real axis, and its limit is the limit of the even convergents of (3). The series (6) converges without a circle of known radius R to this same limit.

The next paragraph contains preliminaries.

2. In the above mentioned article I gave formulas† which may be used to connect the numbers a_i^{-p} of (4) with the a_i^{-p-1} and also with the a_i^{-p+1}. They run as follows:

$$a_{2i}^{-p} = a_{2i+1}^{-p-1} \Big/ \left(\sum_{i=0}^{i-1} a_{2i+1}^{-p-1} \right) \cdot \left(\sum_{i=0}^{i} a_{2i+1}^{-p-1} \right), \tag{8}$$

$$a_{2i-1}^{-p} = a_{2i}^{-p-1} \left(\sum_{i=0}^{i-1} a_{2i+1}^{-p-1} \right)^2, \tag{9}$$

---

* This case was treated in my article, loc. cit., p. 112, Theorem 5. The extensions there obtained were not minimal extensions.

† Wall, loc. cit., formulas (49), (50), (65), (67).

$$a_{2i}^{-p} = a_{2i-1}^{-p+1}\left(c_{-p} - \sum_{i=1}^{i-1} a_{2i}^{-p+1}\right)^2, \tag{10}$$

$$a_{2i+1}^{-p} = a_{2i}^{-p+1} \Big/ \left(c_{-p} - \sum_{i=1}^{i-1} a_{2i}^{-p+1}\right) \cdot \left(c_{-p} - \sum_{i=1}^{i} a_{2i}^{-p+1}\right). \tag{11}$$

If we solve (9) for a_{2i}^{-p-1}, replace p by $p-1$ and equate the value of a_{2i}^{-p} so found to that given by (10) we will obtain, after simple reductions,

$$c_{-p} = \sum_{i=1}^{i-1} a_{2i}^{-p+1} + 1 \Big/ \sum_{i=1}^{i} a_{2i-1}^{-p}. \tag{12}$$

Stieltjes* showed that the sequences of even and odd convergents of the continued fraction

$$\frac{1}{a_1 z} + \frac{1}{a_2} + \frac{1}{a_3 z} + \cdots \tag{13}$$

always converge to limit functions $F_1(z)$ and $F_2(z)$ respectively, and that these limits are expressible as Stieltjes† integrals

$$F_1(z) = \int_0^\infty \frac{d\phi_1(u)}{z+u}, \quad F_2(z) = \int_0^\infty \frac{d\phi_2(u)}{z+u}, \tag{14}$$

where $\phi_1(u)$ and $\phi_2(u)$ are non-decreasing real functions such that $\phi_1(0) = \phi_2(0) = 0$, $\phi_1(\infty) = \phi_2(\infty) = 1/a_1$. The formal expansion of either integral into a power series $P(1/z)$ gives the Stieltjes series corresponding to (13), namely

$$\frac{c_0}{z} - \frac{c_1}{z^2} + \frac{c_2}{z^3} - \cdots, \tag{15}$$

and accordingly $\phi_1(u)$ and $\phi_2(u)$ are functions $\phi(u)$ satisfying the equations

$$\int_0^\infty u^i d\phi(u) = c_i \qquad (i = 0,1,2, \cdots). \tag{16}$$

When $\sum a_i$ diverges, $F_1(z) \equiv F_2(z)$, and all functions $\phi(u)$ satisfying (16) are *equivalent*, i.e. equal at all points of continuity. On the other hand, when $\sum a_i$ converges, $F_1(z) \not\equiv F_2(z)$ and there is an infinite number of non-equivalent functions $\phi(u)$ satisfying (16). In this case the integrals (14) reduce to infinite series of the form

---

* Stieltjes, loc. cit., §§47–48. Note that (13) becomes (3) if we replace z by $1/z$ and then drop the factor z.

† Stieltjes, loc. cit., §38. Cf. also O. Perron, *Die Lehre von den Kettenbrüchen*, 1913, Chapter IX, for the definition and essential properties of Stieltjes integrals, and the chief results of Stieltjes.

$$(17) \qquad F_1(z) = \sum_{i=1}^{\infty} \frac{\mu_i}{z + \lambda_i}, \quad F_2(z) = \frac{\nu_0}{z} + \sum_{i=1}^{\infty} \frac{\nu_i}{z + \theta_i}$$

in which μ_i, λ_i, ν_i, θ_i are all real and positive; and (16) for $\phi(u) = \phi_1(u)$ become

$$(18) \qquad \sum_{i=1}^{\infty} \lambda_i^p \mu_i = c_p \qquad (p = 0,1,2, \cdots),$$

with similar equations for $\phi = \phi_2$.

3. These preliminary remarks having been made, I shall prove the following theorem.

THEOREM 1. *The Stieltjes series* (1) *admits a first extension when and only when the integral*

$$(19) \qquad \int_0^{\infty} \frac{d\phi_1(u)}{u}$$

converges. When this condition is fulfilled we may choose c_{-1} *equal to* (19) *or any greater number.*

For the proof of this theorem the following lemmas will be needed.

LEMMA 1. *If the Stieltjes integrals*

$$\int_0^{\infty} u^k d\phi(u) = c_k \quad (k = 0,1,2, \cdots), \quad \text{and} \quad \phi_1(u) = \int_0^u \frac{d\phi(u)}{u^n},$$

where u is real and positive and n is a positive integer, exist, then $\phi_1(u)$, *which is real, non-negative, and non-decreasing, satisfies the equations*

$$\int_0^{\infty} u^{n+k} d\phi_1(u) = c_k \qquad (k = 0,1,2,3, \cdots).$$

LEMMA 2. *If*

$$\phi_1(u) = \int_0^u u^n d\phi(u),$$

where n is a positive or negative integer or 0, *and* $\phi(u)$ *satisfies the equations*

$$\int_0^{\infty} u^{n+k} d\phi(u) = c_k \qquad (k = 0,1,2,3, \cdots),$$

is convergent, then

$$\int_0^{\infty} u^k d\phi_1(u) = c_k \qquad (k = 0,1,2,3, \cdots).$$

According to the definition of a Stieltjes integral, divide the interval $(0, b)$, $b>0$, in m sub-intervals by the points $(x_0=0<x_1<x_2<\cdots<x_m=b)$, and let the norm of the division be δ. Then if $x_{i-1}\leqq\xi_i\leqq x_i$,

$$\begin{aligned}\int_0^b u^{n+k}d\phi_1(u) &= \lim_{\delta=0}\sum_{i=1}^m \xi_i^{n+k}\left[\int_0^{x_i}\frac{d\phi(u)}{u^n}-\int_0^{x_{i-1}}\frac{d\phi(u)}{u^n}\right]\\ &= \lim_{\delta=0}\sum_{i=1}^m \xi_i^{n+k}\int_{x_{i-1}}^{x_i}\frac{d\phi(u)}{u^n}\\ &= \lim_{\delta=0}\sum_{i=1}^m \xi_i^{n+k}\frac{1}{\xi_i'^{\,n}}[\phi(x_i)-\phi(x_{i-1})],\end{aligned}$$

where ξ_i' is a properly chosen point between x_{i-1} and x_i.* But since $\phi_1(u)$ is a non-decreasing, non-negative, real function, and u^{n+k} is continuous in the interval $(0, b)$, the integral $\int_0^b u^{n+k}d\phi_1(u)$ exists. Consequently we may take $\xi_i=\xi_i'$ and the above limit becomes

$$\int_0^b u^{n+k}d\phi_1(u)=\int_0^b u^k d\phi(u)\qquad(k=0,1,2,\cdots).$$

Now the integral on the right has a limit for $b=\infty$. Hence the integral on the left has a limit for $b=\infty$ and these limits are equal. This proves Lemma 1.

To prove the second lemma, we choose b and $x_0, x_1, x_2, \cdots, x_m$ as above and form the sum

$$(20)\qquad \sum_{i=1}^m \xi_i^k\left[\int_0^{x_i}u^n d\phi(u)-\int_0^{x_{i-1}}u^n d\phi(u)\right]=\sum_{i=1}^m\xi_i^k\int_{x_{i-1}}^{x_i}u^n d\phi(u),$$

which is equal to

$$\sum_{i=1}^m \xi_i^k\xi_i'^n[\phi(x_i)-\phi(x_{i-1})],$$

where ξ_i' is a properly chosen point between x_{i-1} and x_i. But since ξ_i is an arbitrary point in this interval we may take $\xi_i=\xi_i'$. Hence the last sum is equal to

$$\sum_{i=1}^m \xi_i^{n+k}[\phi(x_i)-\phi(x_{i-1})]$$

---

* The theorem here used, which corresponds to the mean value theorem for Riemannian integrals and is proved similarly, is as follows. If $f(x)$ is continuous for $a\leqq x\leqq b$, and $\phi(x)$ is non-decreasing and non-negative then there exists some point ξ, $a\leqq\xi\leqq b$, such that

$$\int_a^b f(x)d\phi(x)=f(\xi)[\phi(b)-\phi(a)].$$

If $f(x)$ is continuous only for $a<x\leqq b$, and $\lim_{x=a}f(x)=+\infty$, the same equation holds with $a<\xi\leqq b$.

which by hypothesis has the limit c_k for $\delta=0$, $b=\infty$. Consequently the left member of (20) has the limit c_k for $\delta=0$, $b=\infty$, and this limit is the integral $\int_0^\infty u^k d\phi_1(u)$. This proves Lemma 2.

We now prove that the condition of Theorem 1 is sufficient for a first extension of (1). Assume that (15) converges and set

$$\phi^{-1}(u) = \int_0^u \frac{d\phi_1(u)}{u}, \qquad \phi^{-1}(0) = 0.$$

Since $\phi_1(u)$ is a solution of (16) we have, by Lemma 1 with $n=1$,

$$\int_0^\infty u^{1+i} d\phi^{-1}(u) = c_i \qquad (i = 0,1,2, \cdots).$$

Thus if

$$\int_0^\infty d\phi^{-1}(u) = \int_0^\infty \frac{d\phi_1(u)}{u} = c_0' \ ; \ c_{i-1} = c_i' \qquad (i = 1,2,3, \cdots),$$

the following equations hold:

$$\int_0^\infty u^i d\phi^{-1}(u) = c_i' \qquad (i = 0,1,2, \cdots).$$

It then follows from the work of Stieltjes that c_0', c_0, c_1, $\cdots$ are coefficients in a Stieltjes series. The sufficiency of the condition is thus proved.

To prove the necessity of the condition, assume that a first extension of (1) exists, and consider separately the cases $\sum a_i$ diverges, $\sum a_i$ converges, respectively.

(a) If $\sum a_i$ diverges, then $c_{-1}=\sum a_{2i}+\delta$, where $\delta \geqq 0$ (§1). If $\delta=0$ it follows from (12) with $p=1$, that $\sum a_{2i-1}^{-1}$ must diverge; and if $\delta>0$, we see from (10) with $p=1$ that $\sum a_{2i}^{-1}$ diverges. Hence in either case $\sum a_i^{-1}$ diverges, and consequently the continued fraction (4) with $p=1$ converges to the limit

$$\frac{1}{z}\int_0^\infty \frac{d\phi^{-1}(u)}{z^{-1}+u}$$

and

$$\int_0^\infty u^{1+i} d\phi^{-1}(u) = c_i \qquad (i = 0,1,2, \cdots).$$

Therefore by Lemma 2 with $n=1$, $\phi(u)=\phi^{-1}(u)$, the function

$$\psi_1(u) = \int_0^u u d\phi^{-1}(u)$$

is a solution of (16), and since $\sum a_i$ diverges this function is equivalent to $\phi_1(u)$.

Let now a, b be real and positive and points of continuity* of $\phi_1(u)$. Then if $b>a$ it follows that

$$\int_a^b \frac{d\phi_1(u)}{u} = \int_a^b \frac{d\psi_1(u)}{u} = \lim_{\delta=0} \sum_{i=1}^{m} \frac{1}{\xi_i'} \cdot \xi_i' [\phi^{-1}(x_i) - \phi^{-1}(x_{i-1})],$$

where ξ_i' is a properly chosen point between x_{i-1} and x_i, $i=1, 2, \cdots, m$, $x_0=a$, $x_m=b$. Thus if $b'>b$,

$$\int_a^{b'} \frac{d\phi_1(u)}{u} = \int_a^b d\phi^{-1}(u) + \int_b^{b'} \frac{d\phi_1(u)}{u}.$$

Now since $\int_0^\infty d\phi_1(u)$ converges, $\int_b^\infty d\phi_1(u)/u$ will surely converge if $b \geqq 1$. Hence for any $\epsilon>0$, there exists a number B such that if $b>B$, $b'>b$,

$$\left| \int_b^{b'} \frac{d\phi_1(u)}{u} \right| < \epsilon,$$

and consequently

$$\lim_{b'=\infty} \int_a^{b'} \frac{d\phi_1(u)}{u} = \int_a^\infty d\phi^{-1}(u) = \phi^{-1}(\infty) - \phi^{-1}(a),$$

or

$$\int_a^\infty \frac{d\phi_1(u)}{u} = c_{-1} - \phi^{-1}(a). \tag{21}$$

If now a approaches 0, over points of continuity of $\phi_1(u)$, the left member of (21) will have the limit†

$$\lim_{a=0^+} \int_a^\infty \frac{d\phi_1(u)}{u} = c_{-1} - \frac{1}{\sum a_{2i-1}^{-1}}. \tag{22}$$

Let a_1 be another point of continuity of $\phi_1(u)$ and let $0<a_1<a'<a$. Then

$$\int_{a'}^\infty \frac{d\phi_1(u)}{u} = \int_{a_1}^\infty \frac{d\phi_1(u)}{u} - \int_{a_1}^{a'} \frac{d\phi_1(u)}{u},$$

or simply

---

* Note that $\phi_1(u)$, being monotone, has points of continuity everywhere dense in the interval $(0, \infty)$.

† Cf. Stieltjes, loc. cit., §58.

$$\int_{a'}^{\infty} = \int_{a_1}^{\infty} - \int_{a_1}^{a'}. \tag{23}$$

Now

$$\int_{a_1}^{a'} = \frac{1}{\xi}\left[\phi_1(a') - \phi_1(a_1)\right], \quad a_1 \leqq \xi \leqq a',$$

and since $\phi_1(u)$ is continuous at a_1 we may make

$$\left| \int_{a_1}^{a'} \right| < \frac{\epsilon}{2}, \quad \text{if } \epsilon > 0,\ a' - a_1 < \delta. \tag{24}$$

Then by (22), (23), (24),

$$\int_{a'}^{\infty} = c_{-1} - 1 \Big/ \sum a_{2i-1}^{-1} + \epsilon, \text{ if } a_1 < \eta,\ a' - a_1 < \delta.$$

Consequently

$$\lim_{a'=0^+} \int_{a'}^{\infty} \frac{d\phi_1(u)}{u} = c_{-1} - 1 \Big/ \sum a_{2i-1}^{-1}.$$

But by (12) with $p=1$, $c_{-1}=\sum a_{2i}+1/\sum a_{2i-1}^{-1}$, and therefore

$$\lim_{a'=0^+} \int_{a'}^{\infty} \frac{d\phi_1(u)}{u} = \int_{0}^{\infty} \frac{d\phi_1(u)}{u} = \sum a_{2i} \leqq c_{-1}.$$

This completes the proof of the theorem for the case that $\sum a_i$ is divergent.

(b) When $\sum a_i$ converges, $\int_0^\infty d\phi_1(u)/(z+u)$ reduces to the first series (17) and therefore if $0<a<\lambda_1$, supposing $\lambda_1<\lambda_2<\cdots$, this integral is equal to $\int_a^\infty d\phi_1(u)/(z+u)$. It then follows by a known theorem* that this integral represents an analytic function for any z not contained in the interval $(-\infty, -a)$. Consequently $\int_0^\infty d\phi_1(u)/u$ converges. Furthermore,

$$\int_0^\infty \frac{d\phi_1(u)}{u} = \sum_{i=1}^{\infty} \frac{\mu_i}{\lambda_i} = \lim_{n=\infty} \frac{P_{2n}(0)}{Q_{2n}(0)} = \sum a_{2i} \leqq c_{-1}$$

inasmuch as $P_{2n}(z)/Q_{2n}(z)$, the $2n$th convergent of (13), has the value $\sum_{i=1}^{n} a_{2i}$ when $z=0$. This completes the proof of Theorem 1.

THEOREM 2. *The Stieltjes series* (1) *admits a minimal kth extension when and only when the integral*

$$\int_0^\infty \frac{d\phi_1(u)}{u^k} \tag{25}$$

converges.

* Perron, loc. cit., p. 369.

Suppose first that (25) converges. Then $\int_0^\infty d\phi_1(u)/u^p$, $p<k$, converges. For if $0<x<x'<\delta<1$,

$$\int_x^{x'} \frac{d\phi_1(u)}{u^p} < \int_x^{x'} \frac{d\phi_1(u)}{u^k} < \epsilon,$$

if δ is sufficiently small.

Taking $p=1$ it follows from Theorem 1 that a first extension exists, and if

$$c_{-1} = \int_0^\infty d\phi_1(u)/u = \sum a_{2i},$$

$\sum a_i^{-1}$ must diverge by (12). Consequently

$$\phi^{-1}(u) = \int_0^u d\phi_1(u)/u.$$

Then taking $p=2$ we find that

$$\int_0^\infty \frac{d\phi^{-1}(u)}{u} = \int_0^\infty \frac{d\phi_1(u)}{u^2} \tag{26}$$

converges and again by Theorem 1, a second extension exists and we take c_{-2} equal to (26), etc. Continuing this argument one will finally arrive at a minimal kth extension of (1).

On the other hand suppose that (1) admits a minimal kth extension, $k\geqq 1$. Then by Theorem 1,

$$\int_0^\infty d\phi_1(u)/u = \sum a_{2i}$$

converges, and c_{-1} has this value. Then by (12) $\sum a_i^{-1}$ diverges and therefore

$$\phi^{-1}(u) = \int_0^u d\phi_1(u)/u.$$

If $k\geqq 2$ it follows from Theorem 1 that

$$\int_0^\infty d\phi^{-1}(u)/u = \int_0^\infty d\phi_1(u)/u^2 = \sum a_{2i}^{-1}$$

converges and is equal to c_{-2}. Hence

$$\phi^{-2}(u) = \int_0^u d\phi_1(u)/u^2,$$

and if $k\geqq 3$,

$$\int_0^\infty d\phi_1(u)/u^3 = \sum a_{2i}^{-2}$$

converges, etc. This argument may evidently be continued until we arrive at the integral $\int_0^\infty d\phi_1(u)/u^k$, whatever value k may have.

4. We next prove the theorem mentioned at the end of §1, namely

THEOREM 3. (a) *If there exist a number $a>0$ such that*

$$\int_0^\infty d\phi_1(u) = \int_a^\infty d\phi_1(u), \tag{27}$$

then (1) *admits a minimal kth extension for all values of k.*

(b) *The continued fraction* (7) *converges to the limit*

$$F_1(z) = \int_0^{1/a} \frac{-ud\phi_1(1/u)}{z+u} \tag{28}$$

which is the limit of the even convergents of (3).

(c) *The series* (6) *converges for all z for which $|z|>1/a$, and represents $F_1(z)$ in that region.*

(d) *In case $\sum a_i$ converges, a may be chosen arbitrarily in the open interval $(0, \lambda_1)$, and $c_{-p}=\sum_{i=1}^\infty \mu_i/\lambda_i^p$, $p=1, 2, 3, \cdots$.*

For by (27)

$$\int_0^\infty d\phi_1(u)/u^k = \int_a^\infty d\phi_1(u)/u^k,$$

and this integral is readily seen to be convergent. Hence, by Theorem 2, (1) admits a minimal kth extension. Consider now the integral (28). We have

$$F_1(z) = \int_0^{1/a} \frac{-ud\phi_1(1/u)}{z+u} = \int_0^{1/a} -u\left[\frac{1}{z} - \frac{u}{z^2} + \frac{u^2}{z^3} - \cdots\right] d\phi_1(1/u).$$

Since the series within the brackets converges uniformly over $(0, 1/a)$ if $|z|>\delta>1/a$, it may be integrated term by term. Therefore

$$\begin{aligned} F_1(z) &= \frac{-\int_0^{1/a} ud\phi_1(1/u)}{z} + \frac{\int_0^{1/a} u^2 d\phi_1(1/u)}{z^2} - \cdots \\ &= \frac{\int_a^\infty d\phi_1(u)/u}{z} - \frac{\int_a^\infty d\phi_1(u)/u^2}{z^2} + \cdots \\ &= \frac{c_{-1}}{z} - \frac{c_{-2}}{z^2} + \cdots, \end{aligned}$$

convergent if $|z|>1/a$. It follows* that the continued fraction (7) converges and is equal to $F_1(z)$. When $\sum a_i$ converges the integrals $\int_0^\infty d\phi_1(u)/u^k$ evidently reduce to the sums $\sum_{i=1}^\infty \mu_i/\lambda_i^k$ by (17).

* Cf. Stieltjes, loc. cit., §10, in which it is shown that when a Stieltjes series converges, the numbers $1/(\alpha_i\alpha_{i+1})$, $i=1, 2, 3, \cdots$, must increase to a finite limit, and consequently $\sum\alpha_i$ must diverge, thus implying the convergence of the continued fraction.

NORTHWESTERN UNIVERSITY,
EVANSTON, ILL.

THE ASYMPTOTIC SOLUTION OF AN OPERATIONAL EQUATION*

BY

JOHN R. CARSON

The subject of this note is the asymptotic solution of the operational equation

$$h = \frac{p^r}{p - \lambda}. \tag{1}$$

For the significance of this equation and general methods of solution, the reader is referred to two papers on the Heaviside operational calculus, which appeared in the Bulletin of the American Mathematical Society; one in the January-February 1926 issue by the writer, and the other in the May-June 1927 issue by March.† For present purposes it is sufficient, however, to state that (1) is a purely symbolic equation in which p symbolically represents the differential operator d/dt, and that $h=h(t)$, for real values of $t \geqq 0$, is uniquely defined and determined either by the Laplace integral equation (valid for all values of p to the right of λ in the complex plane)

$$\int_0^\infty h(t)e^{-pt}dt = \frac{1}{p - \lambda} \cdot \frac{1}{p^{1-r}}, \tag{2}$$

or by the Fourier-Bromwich integral‡

$$h(t) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \frac{e^{tp}}{p - \lambda} \cdot \frac{dp}{p^{1-r}} \qquad (c > \lambda_R). \tag{3}$$

In equation (1), λ is an unrestricted parameter in the complex plane; without loss of essential generality, however, it will be assumed that $|\lambda| = 1$, and $\lambda = e^{i\alpha}$. The exponent r is a proper fraction; thus $0 \leqq r \leqq 1$. For simplicity, r will be assigned the value $\frac{1}{2}$ in the following; the methods of analysis applied to this case are applicable without change to all values of r lying

* Presented to the Society, August 30, 1929; received by the editor of the Bulletin in February, 1929, accepted for publication in the Bulletin, and subsequently transferred to the Transactions; revised copy received by the editors of the Transactions in June, 1929.

† A paper by Wiener, *The operational calculus*, Mathematische Annalen, vol. 95, No. 4, may also be consulted with profit.

‡ Bromwich, Proceedings of the London Mathematical Society, 1916, pp. 410–420. Jeffrey, *Operational Methods in Mathematical Physics.*

between zero and unity. We shall therefore limit explicit consideration to the equation

$$h = \frac{p^{1/2}}{p - \lambda}. \tag{4}$$

The operational equation (4) is of considerable importance in the theory of asymptotic solutions of operational equations, because a fairly large class of operational equations is reducible to sums of expressions of the form of (4). In addition it serves to introduce the question as to whether the divergent solution derived by operational processes is a complete asymptotic expansion, or whether an additional term is required.

In accordance with the usual Heaviside process, a divergent solution of (4) is obtained by the following remarkably simple process. Write (4) as

$$h = -\frac{1}{\lambda}\left(1 - \frac{p}{\lambda}\right)^{-1} p^{1/2},$$

and expand by the binomial theorem without reference to convergence; thus

$$h = -\frac{1}{\lambda}\left(1 + \frac{p}{\lambda} + \left(\frac{p}{\lambda}\right)^2 + \left(\frac{p}{\lambda}\right)^3 + \cdots\right) p^{1/2}. \tag{5}$$

Now replace p^n (n integral) by d^n/dt^n and $p^{1/2}$ by $1/(\pi t)^{1/2}$; (5) thus becomes

$$h \sim -\frac{1}{\lambda(\pi t)^{1/2}}\left\{1 - \frac{1}{2\lambda t} + \frac{1\cdot 3}{(2\lambda t)^2} - \frac{1\cdot 3\cdot 5}{(2\lambda t)^3} + \cdots + (-1)^n \frac{1\cdot 3\cdot 5\cdots(2n-1)}{(2\lambda t)^n} \cdots\right\}. \tag{6}$$

The beautifully simple process sketched above whereby the divergent solution (6) is derived, can be applied to a large class of operational equations in which fractional powers of the operator p appear. Such divergent solutions which are derivable for a large number of operational equations representing physical problems, are true asymptotic solutions. In some cases, however, an extra term must be added, while in most cases the precise asymptotic character of the expansion, and the order of magnitude of the error committed by stopping with the smallest term of the divergent series, require independent investigation. In all cases examined, however, and presumably in all cases, the divergent solution has a real significance, at least for restricted values of the argument t. The analysis of equations (1) and (4) serves as an

excellent illustration of these questions, and is not without interest in its own right.

If we write

$$(7)\quad S_n = -\frac{1}{\lambda(\pi t)^{1/2}}\left\{1 - \frac{1}{2\lambda t} + \frac{1\cdot 3}{(2\lambda t)^2} + \cdots + (-1)^n \frac{1\cdot 3\cdot 5\cdots(2n-1)}{(2\lambda t)^n}\right\},$$

it is easy to show by aid of either (2) or (3) (setting $r = \frac{1}{2}$) that h is equal to S_n plus a remainder. The investigation of the remainder directly from (2) or (3) appears to present difficulties. We can, however, easily derive from (2) the explicit solution

$$(8)\quad h = e^{\lambda t}\int_0^t \frac{e^{-\lambda\tau}}{(\pi\tau)^{1/2}}d\tau.$$

That (8) is the solution of (1) and (2) for $r = \frac{1}{2}$ is easily shown as follows. We have the known identities

$$\int_0^\infty \frac{e^{-pt}}{(\pi t)^{1/2}}dt = \frac{1}{p^{1/2}}$$

and

$$\int_0^\infty e^{\lambda t}\cdot e^{-pt}dt = \frac{1}{p-\lambda}$$

for all values of p to the right of λ in the complex plane. From the foregoing equation (8) follows at once from a theorem due to Borel* which may be stated as follows:

If

$$\int_0^\infty f(t)e^{-pt}dt = F_1(p)\cdot F_2(p),$$

$$\int_0^\infty f_1(t)e^{-pt}dt = F_1(p),$$

$$\int_0^\infty f_2(t)e^{-pt}dt = F_2(p),$$

then

$$f(t) = \int_0^t f_1(t-\tau)f_2(\tau)d\tau = \int_0^t f_1(\tau)f_2(t-\tau)d\tau.$$

---

* *Leçons sur les Séries Divergentes*, 1901, p. 104; see also *Electric Circuit Theory and the Operational Calculus*, New York, McGraw Hill Book Company, p. 41.

Now suppose that the *real part* of $\lambda \geqq 0$. For this case we can write

$$\int_0^t = \int_0^\infty - \int_t^\infty .$$

The first integral on the right hand side is equal to $1/\lambda^{1/2}$, and repeated integration by parts applied to the second integral gives

$$h = \frac{e^{\lambda t}}{\lambda^{1/2}} + S_n + R_n \tag{9}$$

where the remainder term is

$$R_n = (-1)^n \frac{1 \cdot 3 \cdot 5 \cdots (2n+1)}{(2\lambda)^{n+1}} e^{\lambda t} \int_t^\infty \frac{e^{-\lambda \tau}}{\tau^{n+1}(\pi\tau)^{1/2}} d\tau . \tag{10}$$

This remainder term is easily seen to be less than the last term T_n of the divergent series; that is, $|R_n| < |T_n|$. Consequently we see that for the case where the *real part* of $\lambda \geqq 0$, the divergent series (6), directly derived from the operational equation, requires the addition of the extra term $e^{\lambda t}/\lambda^{1/2}$, and that h is asymptotically represented by $e^{\lambda t}/\lambda^{1/2} + S_n$ with an error less than the last term included in the divergent series S_n. This conclusion, it may be remarked, has long been known.

When the *real part* of λ is negative, the problem is quite different, and the method employed above is not applicable. We can, however, write (8) in the form

$$h = \frac{1}{\pi^{1/2}} \int_0^t \frac{e^{\lambda \tau}}{(t-\tau)^{1/2}} d\tau . \tag{11}$$

For this case it has been shown by Gronwall (in an unpublished memorandum communicated to the writer) that S_n is an asymptotic expansion of h. Gronwall, however, did not investigate the error committed by terminating the divergent series, nor did he examine the question as to whether the addition of the "extra term" $e^{\lambda t}/\lambda^{1/2}$ does or does not contribute to numerical accuracy in computation. It is evident, of course, that when S_n is asymptotic to h, $e^{\lambda t}/\lambda^{1/2} + S_n$ is also asymptotic, and vice-versa.

In a recent paper* it is stated that (9) applies for all values of λ (except $\lambda = -1$, in which case the term $e^{\lambda t}/\lambda^{1/2}$ is to be omitted), and that the re-

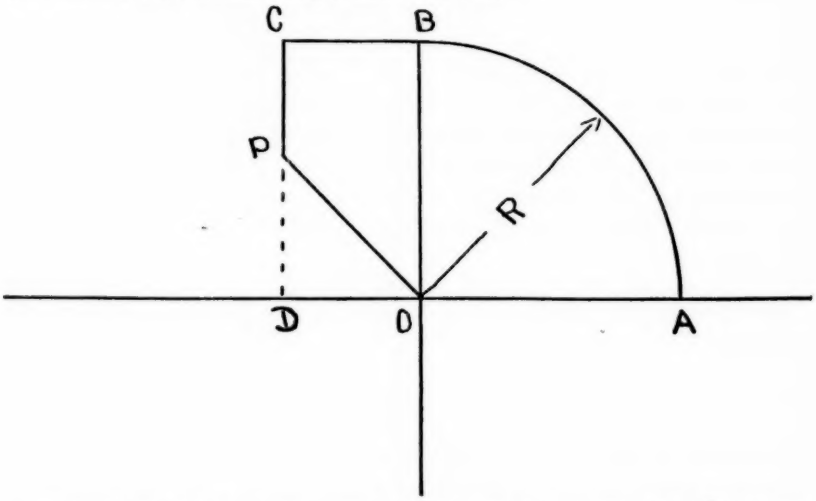
* *Asymptotic expansions of Heaviside's operators*, Berg, Haefner and Smith, Journal of the Franklin Institute, February, 1928. In one and one case only Heaviside himself (*Electromagnetic Theory*, vol. II) adds the extra term when the real part of λ is negative. His justification of his procedure, however, is exceedingly obscure.

mainder term is less than the last term of S_n. This statement is not correct, and the method of analysis employed is invalid. The question, however, admits of simple and rigorous investigation by contour integration, as employed below.

Returning to (8) it may be written as

$$h = \frac{e^{\lambda t}}{(\pi\lambda)^{1/2}} \int_0^{te^{i\alpha}} \frac{e^{-z}}{z^{1/2}} dz \tag{12}$$

representing integration in the complex plane along the straight line $z = e^{i\alpha}$. Since we are concerned with negative values of λ, we shall, without loss of essential generality, assume that λ lies in the second quadrant, and write $\lambda = e^{i\alpha} = -\mu + i\nu$, so that both μ and ν are $\geqq 0$.



Referring to the accompanying sketch, the integral is along the path $0P$ where $P = te^{i\alpha}$. Since the singularity at the branch point $z = 0$ contributes nothing, it follows from Cauchy's theroem that

$$\int_0^P = \int_0^A + \int_A^B + \int_B^C - \int_P^C$$

where AB is the arc of a circle of radius R. If R is made infinite it is easy to see that the second and third integrals vanish, and that

$$\int_0^{P = te^{i\alpha}} = \int_0^{\infty} - \int_{te^{i\alpha}}^{i\infty}.$$

Along the path $0A(A\to\infty)$ set $z=\rho$, where ρ is real. Along the path PC, $z=-\mu t+i\rho$ and the limits of integration are $\rho=\nu t$ and ∞. Consequently (12) becomes

$$h=\frac{e^{\lambda t}}{\lambda^{1/2}}-i\frac{e^{i\nu t}}{(\pi\lambda)^{1/2}}\int_{\nu t}^{\infty}\frac{e^{-i\rho}}{(-\mu t+i\rho)^{1/2}}d\rho. \tag{13}$$

Applying repeated partial integrations to the second integral, we thus get

$$h=\frac{e^{\lambda t}}{\lambda^{1/2}}+S_n+R_n \tag{14}$$

where the remainder term R_n is given by

$$R_n=i(-1)^n\frac{1\cdot3\cdot5\cdots(2n+1)}{2^{n+1}}\cdot\frac{e^{i\nu t}}{(\pi\lambda)^{1/2}}\int_{\nu t}^{\infty}\frac{e^{-i\rho}}{(-\mu t+i\rho)^{n+3/2}}d\rho. \tag{15}$$

If we now denote the extra term $e^{\lambda t}/\lambda^{1/2}$ by T_e; the last retained term of S_n by T_n; and if we write $n+3/2=m$ and $\nu/\mu=\rho$, obvious transformations of (15) lead to the expressions

$$R_n=-\frac{m-1}{\mu^{m-1}}\cdot\frac{1\cdot3\cdot5\cdots(2n-1)}{(2t)^n(\pi t)^{1/2}}e^{\mu t}\cdot I_n\cdot T_e \tag{16}$$

$$=-(m-1)\left(\frac{\lambda}{\mu}\right)^{m-1}e^{i\nu t}I_n\cdot T_n \tag{17}$$

where

$$I_n=\int_{\rho}^{\infty}\frac{e^{i(m\theta-\mu tx)}}{(x^2+1)^{m/2}}dx \tag{18}$$

and

$$\theta=\cos^{-1}\frac{1}{(x^2+1)^{1/2}}.$$

The investigation of the remainder term and the asymptotic character of the expansion is thus reduced to a study of the integral I_n. Equations (16) and (17) conveniently correlate this remainder term with the "extra term" T_e and the last retained term T_n of the divergent series.

Now when a divergent series is employed for numerical computation it is the usual and logical procedure to stop with that term whose absolute value is the numerically smallest. Instead, therefore, of dealing with I_n for unrestricted values of t, we shall assign to t the value n, so that T_n, the last

retained term, is the numerically smallest of the series. This restriction will be understood to apply to the following, and we write

$$I_n = \int_\rho^\infty \frac{e^{i(m\theta - \mu n x)}}{(x^2+1)^{m/2}} dx, \quad m = n + 3/2. \tag{19}$$

If n is made sufficiently large, this integral can be evaluated to any required degree of accuracy by writing $x = \rho + y$, and expanding the integrand in the form

$$\exp\,[-n(c_0 + c_1 y + c_2 y^2 + \cdots)].$$

For sufficiently large values of $t = n$, the integral becomes

$$\frac{1}{2n} e^{(\beta - nc_0)} \left\{ \left(\frac{\pi n}{c_0}\right)^{1/2} - \frac{nc_1}{c_2}\left(1 - \frac{\beta}{3\cdot 1!} + \frac{\beta^2}{5\cdot 2!} \cdots\right)\right\}$$

where $\beta = nc_1^2/(4c_2)$. In this way we find, without going through straightforward details, that for $1 > \rho > 0$,

$$|R_n| \to \frac{T_n}{\{2[1 - (1+\rho^2)^{-1/2}]\}^{1/2}}$$

with increasing values of n. On the other hand, for any fixed large value of n, it is always possible to choose a value of ρ so small that

$$|R_n| = |T_n|\,(\pi n/2)^{1/2}.$$

In the foregoing we have supposed that the index $n = t$ is a large number. In practical applications, however, we are interested in relatively small values of n. For such cases the only way of evaluating the integral I_n is by the pedestrian process of numerical or mechanical quadrature. Fortunately however, the convergence of the integrand is so rapid as to make this process relatively easy and accurate. For the case $n = 5$, $T_n = T_5$ is approximately 1% of the leading term. The results of numerical integration for this case are tabulated below.

ρ	$\lvert R_n\rvert / \lvert T_n\rvert$	$\lvert R_n + T_e\rvert / \lvert R_n\rvert$
0	2.95	0.30
0.1	2.50	0.40
0.2	2.14	0.68
0.3	1.86	1.08
0.4	1.65	1.65
0.5	1.50	2.45
0.8	1.20	6.60
1.0	1.10	10.60

From the foregoing we conclude as follows:

(1) In the region $0 \leqq \rho \leqq 1$, $|R_n|$ is everywhere *greater* than $|T_n|$. The ratio $|R_n|/|T_n|$ increases with decreasing values of ρ and with increasing values of n. In this region, therefore, the error committed by stopping with the term T_n is greater than that term.

(2) If ρ is *less* than a certain value $\rho(n)$, depending on the values of $t=n$ and decreasing therewith, $|T_e+R_n| < |R_n|$. In this region, therefore, the *omission* of the extra term T_e contributes to numerical accuracy. For $n=5$, $\rho(n)=0.3$ approximately, and as n is made indefinitely large, $\rho(n)$ approaches zero. Note, however, that as $n \to \infty$, the value of the extra term becomes negligible compared with S_n.

The foregoing will now be applied to deriving Heaviside's "generalized exponential series."

If the operational equation (4) is expanded in the series

$$h = \frac{1}{p^{1/2}} + \frac{\lambda}{pp^{1/2}} + \frac{\lambda^2}{p^2p^{1/2}} + \cdots,$$

the solution, in accordance with the well known methods of the operational calculus, is then

$$h = 2\left(\frac{t}{\pi}\right)^{1/2}\left\{1 + \frac{1}{3}(2\lambda t) + \frac{1}{3\cdot 5}(2\lambda t)^2 + \cdots\right\}.$$

But it has been shown that

$$h \sim \frac{e^{\lambda t}}{\lambda^{1/2}} + S_n$$

$$\sim \frac{e^{\lambda t}}{\lambda^{1/2}} - \frac{1}{\lambda(\pi t)^{1/2}}\left(1 - \frac{1}{2\lambda t} + \frac{1\cdot 3}{(2\lambda t)^2} - \cdots\right)$$

for all complex values.

Equating we get

$$e^{\lambda t} \sim \sum_{m=-n}^{\infty} \frac{(\lambda t)^{m-1/2}}{\Gamma(m + 1/2)},$$

the expansion for negative values of m being divergent but asymptotic.

More generally, from the operational equation

$$h = \frac{p^r}{p - \lambda} \qquad (0 < r < 1),$$

$$e^{\lambda t} \sim \sum_{m=-n}^{\infty} \frac{(\lambda t)^{m-r}}{\Gamma(m + 1 - r)}.$$

In accordance with the foregoing analysis the expansion ceases to have any significance as λ approaches -1.

This paper will now be closed by a very brief discussion of what I have termed the Heaviside Rule; it is as follows:

If the operational equation $h=1/H(p)$ admits of formal series expansion

$$h = a_0 + a_1 p^{1/2} + a_2 p + a_3 p p^{1/2} + \cdots + a_{2n-1} p^n p^{1/2} + a_{2n} p^n + \cdots,$$

then the equation has a solution, usually divergent and asymptotic,

$$h(t) \sim a_0 + \left(a_1 + a_3 \frac{d}{dt} + a_5 \frac{d^2}{dt^2} + \cdots + a_{2n-1} \frac{d^n}{dt^n} + \cdots\right) \frac{1}{(\pi t)^{1/2}}.$$

To establish this result and the restrictions imposed on its generality we proceed as follows:

The integral equation of the problem

$$\int_0^\infty h(t) e^{-pt} dt = \frac{1}{pH(p)}$$

cannot be employed directly; instead we start with the auxiliary equation defining the auxiliary function $g(t)$:

$$\int_0^\infty g(t) e^{-pt} dt = \frac{1}{p^{n+1} H(p)}.$$

It will be evident at once that $h(t) = (d^n/dt^n) g(t)$. Now we can always write, irrespective of the region of convergence of the expansion,

$$\int_0^\infty g(t) e^{-pt} dt = \frac{1}{p^{n+1}} (a_0 + a_1 p^{1/2} + a_2 p + \cdots + a_{2n-1} p^n p^{1/2} + a_{2n} p^n) + \frac{1}{p^{n+1}} R_n(p),$$

where $R_n(p)$ is the remainder in the series expansion of $1/H(p)$. Remembering that $h = (d^n/dt^n) g$ we have, by well established processes,

$$h(t) = a_0 + \left(a_1 + a_3 \frac{d}{dt} + a_5 \frac{d^2}{dt^2} + \cdots + a_{2n+1} \frac{d^n}{dt^n}\right) \frac{1}{(\pi t)^{1/2}} + r_n(t),$$

where $r_n(t) = (d^n/dt^n) \rho_n(t)$ and ρ_n is defined by the equation

$$\int_0^\infty \rho_n(t) e^{-pt} dt = \frac{1}{p^{n+1}} R_n(p).$$

We observe that by hypothesis $(1/p^{n+1})$ $R_n \sim a_{2n+1}/p^{1/2}$ as $p \to 0$; if we can show correspondingly that $\rho_n(t) \sim a_{2n+1}/(\pi t)^{1/2}$ as $t \to \infty$, then $r_n = O(t^{-(2n+3)/2})$ and the asymptotic character of the expansion is established. The required restrictions on the function $H(p)$ are given by the following proposition:

If in the integral equation

$$\int_0^\infty f(t)e^{-pt}dt = F(p)$$

the function $F(p)$ is everywhere finite and continuous in the right hand side of the complex plane, including the axis of imaginaries, except at $p=0$, in the neighborhood of which $F(p) \sim p^{1/2}$, then $f(t) \sim 1/(\pi t)^{1/2}$ as $t \to \infty$.

In the light of the foregoing, the Heaviside Rule leads to a valid asymptotic solution of the operational equation provided $1/H(p)$ has no singularities on the axis of imaginaries or to the right thereof. Fortunately in the application of the operational calculus to physical problems this condition is satisfied in a great majority of cases, and in the rare cases where it is not, this is usually immediately evident from the physical problem itself. Consequently the engineer and physicist can rely on the Heaviside Rule with confidence.

It will be noted, however, that the foregoing discussion merely establishes the asymptotic character in the sense of Poincaré, and does not determine the magnitude of the error committed by stopping with the smallest term (for finite values of t), nor in particular the effect on numerical accuracy of including "extra" terms corresponding to singularities in $1/H(p)$ to the left of the imaginary axis. Such questions depend, for their answer, upon an independent and usually difficult investigation in each specific case. Furthermore it is a reasonable inference from the analysis in the earlier part of this paper that the inclusion of such extra terms may or may not contribute to numerical accuracy depending on the location of the corresponding singularity in the complex plane. To attempt to include such terms in each case would mean giving up the beautiful simplicity and directness of the Heaviside process for a doubtful gain in numerical accuracy and a very small gain at best in the region where the expansion can be employed with numerical precision.

March (Bulletin of the American Mathematical Society, May-June, 1927) briefly discusses the general problem dealt with above by aid of the Fourier-Bromwich integral (3), and derives the divergent expansion but does not attempt to discuss the magnitude of the error nor to establish the formal asymptotic character of the expansion. Wiener (*The operational calculus*,

Mathematische Annalen, 1926, and appendix to *Operational Circuit Analysis* by V. Bush) discusses the problem in detail and sets up criteria for a true asymptotic expansion. As Wiener himself observes, however, the operational calculus he deals with differs under certain circumstances from that of Heaviside. For a third and quite different treatment of the problem, the reader is referred to a paper by Tibor v. Stacho (*Operatoren Kalkul und Laplacesche Transformation*, publication of the Hungarian University of Francis Joseph, VI, 15, 1927).

American Telephone and Telegraph Company,
New York, N. Y.

ON COMMUTATION FORMULAS IN THE ALGEBRA OF QUANTUM MECHANICS*

BY

NEAL H. McCOY

INTRODUCTION

It is the purpose of this paper to make a study of commutation formulas in the algebra of quantum mechanics. The theories of quantum mechanics introduced by Heisenberg† and Dirac‡ are different in their conception and formulation but both make use of a non-commutative algebra. In Heisenberg's theory, the elements of the algebra are infinite matrices; in Dirac's they are abstract "q-numbers." Schrödinger's§ theory, although mathematically equivalent to Heisenberg's, does not make explicit use of this algebra. However, the operators which Schrödinger uses satisfy the same commutation formulas as Heisenberg's matrices.

We shall consider the algebra of the quantum mechanics from the matric standpoint although the results obtained do not depend upon the form of the variables. The variables enter in pairs as in classical mechanics and we shall use the expression "conjugate quantum variables" or simply "conjugate variables" in analogy with the idea of canonically conjugate variables in the classical theory. For a single pair of conjugate variables the properties of the algebra are determined by the fundamental commutation rule,‖

$$pq - qp = cI, \tag{1}$$

where q and p are matrices representing the coördinate and momentum respectively, c is a real or complex number and I is the unit matrix. In the quantum mechanics $c = h/(2\pi i)$, although the algebra does not depend upon the particular value assigned to c. The symbol I will be omitted but will be understood wherever a real or complex number occurs alone in a matric equation.

---

* Presented to the Society March 30, 1929; received by the editors in June, 1929.

† Heisenberg, W., Zeitschrift für Physik, vol. 33 (1925), pp. 879–893; Born and Jordan, Zeitschrift für Physik, vol. 34 (1925), pp. 858–888.

‡ Dirac, P. A. M., Proceedings of the Royal Society, (A), vol. 109 (1925), pp. 642–653.

§ Various papers in the Annalen der Physik beginning with vol. 79 (1926), pp. 361–376.

‖ Born and Jordan, loc. cit., p. 871.

For more than a single pair of conjugate quantum variables the relation (1) is replaced by the following:*

$$p_r q_s - q_s p_r = c\delta_{rs},$$
$$p_r p_s - p_s p_r = 0, \tag{2}$$
$$q_r q_s - q_s q_r = 0.$$

We shall be interested primarily in polynomials in a certain number of these variables. The results obtained in this paper are rigorously established for polynomials only, although they are formally correct for infinite series and thus give valid results if the series are convergent.

Let f and g be any two polynomials in the conjugate quantum variables, p and q, subject to the condition (1). In general $fg \neq gf$, but the relation (1) makes it possible to compute $fg-gf$ when f and g are given. This type of calculation arises frequently as, for example, in taking time derivatives. In order to simplify the calculations, the following formulas were derived:†

$$pg - gp = c\frac{\partial g}{\partial q}, \tag{3}$$

$$gq - qg = c\frac{\partial g}{\partial p}, \tag{4}$$

where g is a polynomial in p and q. We shall find it convenient to consider these equations as defining differentiation. It follows that the usual formulas for differentiating polynomials hold.‡

The obvious advantages of the formulas (3) and (4) suggest the desirability of obtaining a formula for computing $fg-gf$ directly, where f and g are arbitrary polynomials. Such a formula has been found. This formula reduces to equations (3) or (4) if one of the functions is taken as p or q, and it becomes identical with condition (1) if $f=p$, $g=q$.

This general commutation formula has been extended to functions of any number of quantum variables and applied to some special problems.§

---

* Dirac, loc. cit., p. 649; Born, Heisenberg and Jordan, Zeitschrift für Physik, vol. 35 (1926), p. 573.

† Born, Heisenberg and Jordan, loc. cit., p. 563.

‡ Care must be taken to preserve the order of factors, that is,

$$\frac{\partial(fg)}{\partial q} = f\frac{\partial g}{\partial q} + \frac{\partial f}{\partial q}g.$$

§ A summary of some of the results obtained has been published in Proceedings of the National Academy of Sciences, vol. 15, pp. 200–202, March, 1929.

In the second part of this paper we extend these commutation formulas to the case of vectors in quantum mechanics. Born and Jordan* introduced the idea of using vectors whose components are functions of the quantum variables and Pauli† proved its usefulness in the theory of the hydrogen atom.

Let $A=(A_1, A_2, A_3)$ and $B=(B_1, B_2, B_3)$ be any two vectors. Scalar and vector products are defined by

$$A\cdot B = A_1B_1 + A_2B_2 + A_3B_3,$$

and

$$A \times B = (A_2B_3 - A_3B_2, A_3B_1 - A_1B_3, A_1B_2 - A_2B_1).$$

These are the ordinary definitions except that here the order A, B must be preserved throughout. It follows that $A\cdot B \neq B\cdot A$ and $A\times B \neq -B\times A$. This constitutes a difference from ordinary vector analysis. General commutation formulas for these scalar and vector products have been obtained.

I. Commutation formulas in quantum algebra

1. **Transformations of identities.**‡ We shall establish two theorems on transformations of identities in the single pair of quantum variables, p and q.

Theorem I. *Any identity remains an identity if p, q, c are replaced by p', q', Δc respectively, where p', q' are obtained from p, q by the non-singular linear transformation*

$$p' = \alpha p + \beta q,$$
$$q' = \gamma p + \delta q$$

of determinant Δ, the coefficients α, β, γ, δ being real or complex numbers.

This theorem is readily established by noting that $p'q'-q'p'=\Delta(pq-qp)=\Delta c$. Thus p', q' have an algebra equivalent to that of p, q with c replaced by Δc. This theorem was stated by Dirac§ for the special transformation given by $\alpha=\delta=0, \beta=\gamma=1$, which is equivalent to the interchange of p and q.

The following theorem was also given by Dirac‖ but without detailed proof.

Theorem II. *Any identity remains an identity if the order of all factors is reversed and c replaced by $-c$.*

---

* Loc. cit.

† Zeitschrift für Physik, vol. 36 (1926), pp. 336–363.

‡ The word "equation" will be frequently used to apply to an identity. We shall consider no equations of condition among the quantum variables.

§ Dirac, Proceedings of the Royal Society, (A), vol. 110 (1926), p. 565.

‖ Ibid.

The meaning of this theorem will be made clear by an example. Consider the identity

$$\frac{1}{pq+c}qp = \frac{1}{qp+2c}(qp-c).$$

According to the theorem,

$$pq\frac{1}{qp-c} = (qp+c)\frac{1}{pq-2c},$$

which may easily be verified by reducing it to the form

$$pq\frac{1}{qp-c} = pq\frac{1}{qp-c}.$$

An identity of this last type which merely states the equivalence of two expressions of the same form will be called a formal identity.

Any identity may be reduced to a formal identity by means of the following transformations:

(a) $$pq = qp + c,$$

(b) $$qp = pq - c,$$

together with algebraic reduction. By reversing this process it follows that any identity may be generated from a formal identity by using these same transformations. We shall consider identities which are thus obtained from formal identities in a finite number of steps.

Theorem II is evident for formal identities. Hence we have only to show that if the theorem is true for a given identity it is true for those obtained from it by the transformations given above.

If ϕ is any expression in p and q we shall use $\bar{\phi}$ to denote the expression obtained from ϕ by reversing the order of all factors and changing the sign of c. Let $f=0$ be any identity for which $\bar{f}=0$ also. If we replace any factor of the form pq in a single term of the expression f by $qp+c$, the result will be an identity which will be denoted by $g=0$. The expression g differs from f in only one term and there only in that pq has been replaced by $qp+c$. Thus $\bar{g}$ differs from $\bar{f}$ only in that qp has taken the place of $pq-c$ and it follows that $\bar{g}=0$. The argument follows in a similar manner if a factor of the form qp is replaced by $pq-c$. The proof is completed by noting that if the theorem is true for a given identity it is true for any identity obtained from this one by algebraic manipulation.

Theorems I and II are extended without difficulty to identities in more than one pair of variables.

2. A general commutation formula for functions of one pair of quantum variables. We shall prove the following theorem:

THEOREM III. *If f and g are arbitrary polynomials in p and q, then* (a)

$$fg - gf = \sum_{s=1} \frac{c^s}{s!}\left[\frac{\partial^s g}{\partial q^s}\frac{\partial^s f}{\partial p^s} - \frac{\partial^s f}{\partial q^s}\frac{\partial^s g}{\partial p^s}\right], \tag{5}$$

and (b)

$$fg - gf = \sum_{s=1} \frac{(-c)^s}{s!}\left[\frac{\partial^s g}{\partial p^s}\frac{\partial^s f}{\partial q^s} - \frac{\partial^s f}{\partial p^s}\frac{\partial^s g}{\partial q^s}\right], \tag{6}$$

the sum in each case being taken over all non-null derivatives of f and g.

It will be seen that formulas (5), (6) reduce to equations (3), (4) in the special cases $f=p, f=q$. While the theorem is stated for polynomials only it is formally correct for other analytic functions. It will give valid results if one of the functions is a polynomial and the other is any function whose first and higher derivatives exist under the definition given by equations (3) and (4).

An outline of the method of proof of the first part of this theorem is as follows.* The validity of formula (5) in the special cases $f=p, f=q$ has already been noted. We shall show later that if formula (5) is true for a given polynomial f with g an arbitrary polynomial it remains so if f is replaced by pf or by qf. Since any term of a polynomial may be obtained by starting with either p or q and multiplying by p or q on the left in the proper order, it follows that equation (5) is true if f is any term of a polynomial. Furthermore this equation is valid for the sum of two functions if it is for each separately and thus the function f may be any polynomial.

For convenience we shall introduce the notation

$$H^{(s)}(g,f) = \frac{\partial^s g}{\partial q^s}\frac{\partial^s f}{\partial p^s}.$$

Equation (5) may then be written in the form

$$fg - gf = \sum_{s=1} \frac{c^s}{s!}[H^{(s)}(g,f) - H^{(s)}(f,g)]. \tag{5'}$$

---

* Another proof of formula (5) where f and g are arbitrary functions subject to the condition that they be in the normal form for quantum mechanics has been given by M. Coulomb, Comptes Rendus, vol. 188, No. 20, May 13, 1929. Obviously no such restriction is made in the proof given here.

The following auxiliary relations may now be verified:

$$H^{(s)}(qf,g) = qH^{(s)}(f,g) + sH^{(s-1)}\left(f,\frac{\partial g}{\partial p}\right); \tag{7}$$

$$H^{(s)}(g,qf) = qH^{(s)}(g,f) + cH^{(s)}\left(\frac{\partial g}{\partial p},f\right); \tag{8}$$

$$H^{(s)}(pf,g) = pH^{(s)}(f,g); \tag{9}$$

$$H^{(s)}(g,pf) = pH^{(s)}(g,f) - cH^{(s)}\left(\frac{\partial g}{\partial q},f\right) + sH^{(s-1)}\left(\frac{\partial g}{\partial q},f\right). \tag{10}$$

Consider the last of these relations. Making use of the fact that

$$\frac{\partial^s(pf)}{\partial p^s} = p\frac{\partial^s f}{\partial p^s} + s\frac{\partial^{s-1}f}{\partial p^{s-1}},$$

we have

$$\begin{aligned} H^{(s)}(g,pf) &= \frac{\partial^s g}{\partial q^s}p\frac{\partial^s f}{\partial p^s} + s\frac{\partial^s g}{\partial q^s}\frac{\partial^{s-1}f}{\partial p^{s-1}} \\ &= p\frac{\partial^s g}{\partial q^s}\frac{\partial^s f}{\partial p^s} - c\frac{\partial^{s+1}g}{\partial q^{s+1}}\frac{\partial^s f}{\partial p^s} + s\frac{\partial^s g}{\partial q^s}\frac{\partial^{s-1}f}{\partial p^{s-1}} \end{aligned}$$

by equation (3). This is equation (10) in a different notation. The other relations may be verified in a similar manner.

We now assume that the function f satisfies equation (5′) with g an arbitrary polynomial, and shall consider the effect of replacing f by pf. We wish to show that

$$pfg - gpf = \sum_{s=1} \frac{c^s}{s!}[H^{(s)}(g,pf) - H^{(s)}(pf,g)].$$

By using relations (9), (10), this may be written in the form

$$pfg - gpf = \sum_{s=1} \frac{c^s}{s!}\left[pH^{(s)}(g,f) - cH^{(s)}\left(\frac{\partial g}{\partial q},f\right) + sH^{(s-1)}\left(\frac{\partial g}{\partial q},f\right) - pH^{(s)}(f,g)\right]. \tag{11}$$

In order to verify this equation, multiply equation (3) by f on the right and equation (5′) by p on the left and add. We get the result

$$(12) \qquad pfg - gpf = p\sum_{s=1}\frac{c^s}{s!}[H^{(s)}(g,f) - H^{(s)}(f,g)] + c\frac{\partial g}{\partial q}f.$$

Subtract equation (11) from equation (12). We obtain

$$c\left\{\sum_{s=1}\frac{c^s}{s!}H^{(s)}\left(\frac{\partial g}{\partial q},f\right) - \sum_{s=1}\frac{c^{s-1}}{(s-1)!}H^{(s-1)}\left(\frac{\partial g}{\partial q},f\right) + \frac{\partial g}{\partial q}f\right\}$$

$$= c\left\{\sum_{s=1}\frac{c^s}{s!}H^{(s)}\left(\frac{\partial g}{\partial q},f\right) - \sum_{s=1}\frac{c^s}{s!}H^{(s)}\left(\frac{\partial g}{\partial q},f\right) - \frac{\partial g}{\partial q}f + \frac{\partial g}{\partial q}f\right\} = 0,$$

which establishes the equivalence of these equations. Since equation (12) is known to be correct, we have thus established the validity of equation (11).

The case f is replaced by qf may be disposed of similarly except for the necessity of showing that

$$c\left\{f\frac{\partial g}{\partial p} - \frac{\partial g}{\partial p}f - \sum_{s=1}\frac{c^s}{s!}\left[H^{(s)}\left(\frac{\partial g}{\partial p},f\right) - H^{(s)}\left(f,\frac{\partial g}{\partial p}\right)\right]\right\} = 0.$$

This is true since the polynomial g in (5′) is arbitrary and thus may be replaced by $\partial g/\partial p$. This completes the proof of the first part of Theorem III.

The proof of the second part is as follows. Let $f=f(p, q)$, $g=g(p, q)$ be the functions considered. By equation (5) and Theorem I we know that

$$(13) \quad f(q,p)g(q,p) - g(q,p)f(q,p) = \sum_{s=1}\frac{(-c)^s}{s!}\left[\frac{\partial^s g(q,p)}{\partial p^s}\frac{\partial^s f(q,p)}{\partial q^s} - \frac{\partial^s f(q,p)}{\partial p^s}\frac{\partial^s g(q,p)}{\partial q^s}\right].$$

Since $f(p, q)$, $g(p, q)$ are arbitrary polynomials it follows that $f(q, p)$, $g(q, p)$ are also arbitrary polynomials. If we replace $f(q, p)$, $g(q, p)$ by $f(p, q)$, $g(p, q)$ in equation (13) we are led at once to formula (6).

3. **An application to expressions of the type $q^m p^{m'} q^n p^{n'}$.** As an important special case of formula (5) let $f=p^n$, $g=q^m$. We find

$$p^n q^m - q^m p^n = \sum_{s=1} c^s s!\binom{n}{s}\binom{m}{s} q^{m-s}p^{n-s}$$

or

$$(14) \qquad p^n q^m = \sum_{s=0} c^s s!\binom{n}{s}\binom{m}{s} q^{m-s}p^{n-s}.$$

We shall interpret $\binom{a}{b}$ to be zero if $a<b$. The corresponding expression for $q^m p^n$ may be obtained by formula (6). The result is

$$(15) \qquad q^m p^n = \sum_{s=0}(-c)^s s!\binom{n}{s}\binom{m}{s} p^{n-s}q^{m-s}.$$

The relations (14) and (15) were given by Born and Jordan.*

Let us consider the expansion of $q^m p^{m'} q^n p^{n'}$ in the form $\sum a_{ij} p^i q^j$. We get, by an application of (15),

$$q^m p^{m'} q^n p^{n'} = \sum_{s=0} (-c)^s s! \binom{m}{s} \binom{m'}{s} p^{m'-s} q^{m+n-s} p^{n'}.$$

Repeating the process with $q^{m+n-s} p^{n'}$, we find

$$(16)\quad q^m p^{m'} q^n p^{n'} = \sum_{s=0} \sum_{r=0} (-c)^{s+r} s! r! \binom{m'}{s} \binom{n'}{r} \binom{m}{s} \binom{m+n-s}{r} p^{m'+n'-s-r} q^{m+n-s-r}.$$

By making certain restrictions on the exponents, m, m', n, n', we may obtain some interesting results. Assume that the left member of equation (16) is isobaric in p and q, that is, $m+n=m'+n'$. Reverse the order of all factors, at the same time interchanging p and q. Since the right member is unchanged, while the left member becomes $q^{n'} p^n q^{m'} p^m$, we have

$$q^m p^{m'} q^n p^{n'} = q^{n'} p^n q^{m'} p^m.$$

It will be seen that in general any term isobaric in p and q will have an expansion similar to (16) which is unchanged by this transformation. We have then the following theorem.

THEOREM IV. *Let g be any term isobaric in p and q. Interchange p and q and reverse the order of all factors in g. If $\bar{g}$ denotes the resulting term, then $g=\bar{g}$.*

Since $q^m p^{m'} q^n p^{n'} = q^{n'} p^n q^{m'} p^m$, the coefficient of $p^{m+n-k} q^{m+n-k}$ must be the same in their expansions, and we are led to the identity

$$(17)\quad \sum_{s+r=k} s! r! \binom{m'}{s} \binom{n'}{r} \binom{m}{s} \binom{m+n-s}{r} = \sum_{s+r=k} s! r! \binom{n}{s} \binom{m}{r} \binom{n'}{s} \binom{m+n-s}{r}.$$

If we divide by $m!m'!n!n'!\{k![(m+n-k)!]^2\}^{-1}$ this may be written in the form

$$(17')\quad \sum_{s=0} \binom{k}{s} \binom{m+n-s}{n} \binom{m+n-k}{m'-s} = \sum_{s=0} \binom{k}{s} \binom{m+n-s}{m'} \binom{m+n-k}{n-s}.$$

This identity may be proved directly by induction.

* Loc. cit., p. 873.

We may get a series of identities in a similar manner. If we consider the expansion of $q^m p^{m'} q^n p^{n'} q^l p^{l'}$ under the condition $m+n+l=m'+n'+l'$ we are led as above to the identity

$$\sum_{s+r+t=k} \binom{k}{s}\binom{k-s}{r}\binom{m+n-s}{n}\binom{m'+n'-s-r}{n'-r}\binom{m+n+l-s-r}{l}\binom{m+n+l-k}{l'-t}$$

(18)

$$= \sum_{s+r+t=k} \binom{k}{s}\binom{k-s}{r}\binom{l'+n'-s}{n'}\binom{l+n-s-r}{n-r}\binom{m+n+l-s-r}{m'}\binom{m+n+l-k}{m-t}.$$

In the special case $n=n'=0$, it follows that $l+m=l'+m'$ and $r=0$. The resulting identity may be written in the form of (17′) with l replacing n.

4. **Taylor's series and the commutation formula.** It may be shown for polynomials without difficulty that

$$f(p+h,q+k) = \sum_{s=0} \sum_{\alpha=0}^{s} \frac{h^\alpha k^{s-\alpha}}{\alpha!(s-\alpha)!} \frac{\partial^s f(p,q)}{\partial p^\alpha \partial q^{s-\alpha}} \tag{19}$$

where h and k are real or complex numbers. We thus have a Taylor's series expansion as in the case of functions of a real variable. For analytic functions other than polynomials, equation (19) is formally correct.

Some interesting formal results may be obtained from this expansion and the commutation formula (5). From (5) we have

$$f(p,q)e^{nq} - e^{nq}f(p,q) = e^{nq}\sum_{s=1} \frac{c^s n^s}{s!} \frac{\partial^s f(p,q)}{\partial p^s} = e^{nq}[f(p+nc,q) - f(p,q)]$$

by equation (19). From this it follows that

$$\begin{aligned} f(p,q)e^{nq} &= e^{nq}f(p+nc,q), \\ e^{nq}f(p,q) &= f(p-nc,q)e^{nq}. \end{aligned} \tag{20}$$

These results were first given by Dirac* and used in his theory of the hydrogen atom.

In the special case $f(p,q)=e^{p}$, $n=1$, the first of equations (20) reduces to

$$e^p e^q = e^c e^q e^p.$$

This result may also be obtained directly from relation (5) without the use of Taylor's series by placing $f=e^p$, $g=e^q$.

In a similar manner it may be shown that

---

* Dirac, Proceedings of the Royal Society, (A), vol. 110 (1926), p. 566.

$$f(p,q)e^{p+q} - e^{p+q}f(p,q) = \sum_{s=1} \frac{c^s}{s!}\left[e^{p+q}\frac{\partial^s f(p,q)}{\partial p^s} - \frac{\partial^s f(p,q)}{\partial q^s}e^{p+q}\right]$$
$$= e^{p+q}[f(p+c,q) - f(p,q)] - [f(p,q+c) - f(p,q)]e^{p+q},$$

from which we get the interesting result

$$e^{p+q}f(p+c,q) = f(p,q+c)e^{p+q}.$$

5. **Extension of the general commutation formula to functions of n pairs of variables.** In the case of n pairs of conjugate quantum variables, the single commutation relation (1) is replaced by the set of relations (2). Equations (3) and (4) are replaced by the following:†

$$p_r g - g p_r = c\frac{\partial g}{\partial q_r}, \quad g q_r - q_r g = c\frac{\partial g}{\partial p_r}, \tag{21}$$

where now g is a function of all the variables.

The extensions of the general commutation formulas (5) and (6) are made by the following theorem.

THEOREM V. *If f and g are polynomials in n pairs of variables, then* (a)

$$fg - gf = \sum_{s=1} \frac{c^s}{s!}[H^{(s)}(g,f) - H^{(s)}(f,g)] \tag{22}$$

where

$$H^{(s)}(g,f) = \sum_{\alpha+\beta+\cdots+\delta=s} \frac{s!}{\alpha!\beta!\cdots\delta!}\,\frac{\partial^s g}{\partial q_1^{\alpha}\partial q_2^{\beta}\cdots\partial q_n^{\delta}}\,\frac{\partial^s f}{\partial p_1^{\alpha}\partial p_2^{\beta}\cdots\partial p_n^{\delta}},$$

and (b)

$$fg - gf = \sum_{s=1} \frac{(-c)^s}{s!}[H^{*(s)}(g,f) - H^{*(s)}(f,g)] \tag{23}$$

where

$$H^{*(s)}(g,f) = \sum_{\alpha+\beta+\cdots+\delta=s} \frac{s!}{\alpha!\beta!\cdots\delta!}\,\frac{\partial^s g}{\partial p_1^{\alpha}\partial p_2^{\beta}\cdots\partial p_n^{\delta}}\,\frac{\partial^s f}{\partial q_1^{\alpha}\partial q_2^{\beta}\cdots\partial q_n^{\delta}}.$$

The following relations hold as in the case of one pair of variables:

$$H^{(s)}(q_1 f,g) = q_1 H^{(s)}(f,g) + sH^{(s-1)}\left(f,\frac{\partial g}{\partial p_1}\right),$$

$$H^{(s)}(g,q_1 f) = q_1 H^{(s)}(g,f) + cH^{(s)}\left(\frac{\partial g}{\partial p_1},f\right),$$

$$H^{(s)}(p_1 f,g) = p_1 H^{(s)}(f,g),$$

$$H^{(s)}(g,p_1 f) = p_1 H^{(s)}(g,f) - cH^{(s)}\left(\frac{\partial g}{\partial q_1},f\right) + sH^{(s-1)}\left(\frac{\partial g}{\partial q_1},f\right).$$

† Born, Heisenberg and Jordan, loc. cit., p. 574.

Consider the proof of the last of these expressions. We find

$$H^{(s)}(g,p_1f) = \sum_{\alpha+\beta+\cdots+\delta=s} \frac{s!}{\alpha!\beta!\cdots\delta!} \frac{\partial^s g}{\partial q_1^\alpha \partial q_2^\beta \cdots \partial q_n^\delta} p_1 \frac{\partial^s f}{\partial p_1^\alpha \partial p_2 \cdots \partial p_n^\delta}$$

$$+ \sum_{\alpha+\beta+\cdots+\delta=s} \frac{s!\alpha}{\alpha!\beta!\cdots\delta!} \frac{\partial^s g}{\partial q_1^\alpha \partial q_2^\beta \cdots \partial q_n^\delta} \frac{\partial^{s-1} f}{\partial p_1^{\alpha-1} \partial p_2^\beta \cdots \partial p_n^\delta}$$

$$= p_1 \sum_{\alpha+\beta+\cdots+\delta=s} \frac{s!}{\alpha!\beta!\cdots\delta!} \frac{\partial^s g}{\partial q_1^\alpha \partial q_2^\beta \cdots \partial q_n^\delta} \frac{\partial^s f}{\partial p_1^\alpha \partial p_2^\beta \cdots \partial p_n^\delta}$$

$$- c \sum_{\alpha+\beta+\cdots+\delta=s} \frac{s!}{\alpha!\beta!\cdots\delta!} \frac{\partial^{s+1} g}{\partial q_1^{\alpha+1} \partial q_2^\beta \cdots \partial q_n^\delta} \frac{\partial^s f}{\partial p_1^\alpha \partial p_2^\beta \cdots \partial p_n^\delta}$$

$$+ s \sum_{\alpha+\beta+\cdots+\delta=s} \frac{(s-1)!}{(\alpha-1)!\beta!\cdots\delta!} \frac{\partial^s g}{\partial q_1^\alpha \partial q_2^\beta \cdots \partial q_n^\delta} \frac{\partial^{s-1} f}{\partial p_1^{\alpha-1} \partial p_2^\beta \cdots \partial p_n^\delta}$$

by equations (21). This is the desired result. The other relations may be verified in a similar manner.

Relation (22) may now be established as follows. Assume it to be true for polynomials in n pairs of variables. It remains valid if g is a polynomial in $n+1$ pairs of variables and f is a polynomial in n pairs of variables. Suppose, for convenience, that f is a polynomial in the variables $(q_2, q_3, \cdots, q_{n+1}, p_2, p_3, \cdots, p_{n+1})$. It may be shown as in the proof of Theorem III(a) that equation (22) is still valid if f is replaced by $p_1 f$ or by $q_1 f$. It follows that the relation (22) is true for polynomials in $n+1$ pairs of variables. The case $n=1$ is disposed of directly by Theorem III and thus formula (22) is true for polynomials in any number of variables.

The proof of Theorem V(b) follows from Theorem V(a) and Theorem I as in the proof of Theorem III(b).

The expression for $H^{(s)}(g,f)$ assumes a rather simple form in the case of two pairs of variables. It may be written in the form

$$H^{(s)}(g,f) = \sum_{\alpha=0}^{s} \frac{s!}{\alpha!(s-\alpha)!} \frac{\partial^s g}{\partial q_1^\alpha \partial q_2^{s-\alpha}} \frac{\partial^s f}{\partial p_1^\alpha \partial p_2^{s-\alpha}}.$$

This is seen to be a symbolic expansion of

$$\left(\frac{\partial g}{\partial q_1}\frac{\partial f}{\partial p_1} + \frac{\partial g}{\partial q_2}\frac{\partial f}{\partial p_2}\right)^s.$$

The corresponding definitions of $H^{(s)}(g,f)$ in the case of more variables are the corresponding multinomial expansions.

6. **Extension of the theory to rational functions of the quantum variables.** We shall now discuss a few commutation problems involving rational functions of p and q. The existence of these functions is assumed.

From equations (3) and (4) it follows that

$$\frac{\partial(1/f)}{\partial p} = -\frac{1}{f}\frac{\partial f}{\partial p}\frac{1}{f},$$

if f is any polynomial. If $f=p$, we see that $\partial(1/p)/\partial p = -1/p^2$ as usual.

Let us consider the possibility of extending the general commutation formula (5) to rational functions. In the special case $f=1/p$, $g=1/q$ we find the formal result

$$\frac{1}{p}\frac{1}{q} - \frac{1}{q}\frac{1}{p} = \sum_{s=1}^{\infty} c^s s! \frac{1}{q^{s+1}}\frac{1}{p^{s+1}},$$

which is obviously divergent.

Suppose, however, that f is a polynomial while g is an arbitrary rational function of p and q. The proof of formula (5) is seen to include such cases. Since f is a polynomial we are assured of the termination of the right hand member. As f and g play similar rôles it is clear that we may also take f as the more general function if g is a polynomial. This argument applies also to the commutation formulas for functions of n pairs of variables.

An interesting special case arises when f is a polynomial and $g=1/q$. Then

$$f\frac{1}{q} - \frac{1}{q}f = \sum_{s=1}(-c)^s \frac{1}{q^{s+1}}\frac{\partial^s f}{\partial p^s}.$$

While the general commutation formulas do not give useful results when both f and g involve negative powers of p and q, a given expression can often be transformed into one to which the formulas can be applied. As an example let us calculate $(1/p^m)(1/q^n)-(1/q^n)(1/p^m)$. If this expression is represented by X it follows that

$$p^m X p^m = \frac{1}{q^n}p^m - p^m\frac{1}{q^n} = -\sum_{s=1}(-c)^s s!\binom{n+s-1}{s}\binom{m}{s}\frac{1}{q^{n+s}}p^{m-s}$$

or

$$X = -\sum_{s=1}(-c)^s s!\binom{n+s-1}{s}\binom{m}{s}\frac{1}{p^m}\frac{1}{q^{n+s}}\frac{1}{p^s}.$$

From this we may write the identity

$$\frac{1}{q^n}\frac{1}{p^m} = \sum_{s=0}(-c)^s s!\binom{n+s-1}{s}\binom{m}{s}\frac{1}{p^m}\frac{1}{q^{n+s}}\frac{1}{p^s}.$$

A different expansion may be obtained by considering $q^n X q^n$. We obtain in this way

$$\frac{1}{q^n}\frac{1}{p^m} = \sum_{s=0}(-c)^s s!\binom{m+s-1}{s}\binom{n}{s}\frac{1}{q^s}\frac{1}{p^{m+s}}\frac{1}{q^n}.$$

II. Commutation rules in vector analysis

7. **Introduction of vector notation.** In this part we shall consider vectors whose components are functions of one or more pairs of the quantum variables discussed previously.

We shall let f, g denote scalar functions and F, G vector functions of the quantum variables. We define

$$\text{grad}_Q f = \left(\frac{\partial f}{\partial q_1}, \frac{\partial f}{\partial q_2}, \frac{\partial f}{\partial q_3}\right),$$

$$\text{grad}_P f = \left(\frac{\partial f}{\partial p_1}, \frac{\partial f}{\partial p_2}, \frac{\partial f}{\partial p_3}\right).$$

Corresponding definitions will be assumed for div F, curl F, $\nabla^2 F$.

It is found that a large number of relations go over unchanged from the classical theory.* We shall consider some commutation formulas which of course do not ordinarily appear.

8. **Fundamental commutation relations.** Consider the two vectors $P=(p_1, p_2, p_3)$ and $Q=(q_1, q_2, q_3)$ where the components satisfy relations (2). It follows that

$$P\cdot Q - Q\cdot P = 3c,$$

and

$$P\times Q + Q\times P = 0.$$

Corresponding to previous relations of the type

$$p_r f - f p_r = c\frac{\partial f}{\partial q_r},$$

we find the following formulas which may be easily verified:

$$\begin{aligned}
&\text{(a)} && Pf - fP = c\,\text{grad}_Q f,\\
&\text{(b)} && fQ - Qf = c\,\text{grad}_P f,\\
&\text{(c)} && P\cdot F - F\cdot P = c\,\text{div}_Q F,\\
&\text{(d)} && F\cdot Q - Q\cdot F = c\,\text{div}_P F,\\
&\text{(e)} && F\times P + P\times F = c\,\text{curl}_Q F,\\
&\text{(f)} && F\times Q + Q\times F = -c\,\text{curl}_P F.
\end{aligned} \tag{24}$$

* Pauli, loc. cit.; Handbuch der Physik, Band XX, *Licht als Wellenbewegung*, Berlin, Springer, 1928.

9. **General commutation formulas.** The general commutation formulas developed above may be extended without difficulty to vector functions. If we define

$$\frac{\partial F}{\partial q_1} = \left(\frac{\partial F_1}{\partial q_1}, \frac{\partial F_2}{\partial q_1}, \frac{\partial F_3}{\partial q_1}\right),$$

Theorem V is still applicable if one of the functions is a scalar and the other a vector.

In the case of scalar and vector products we have the following theorem, the proof of which follows without difficulty from Theorem V.

THEOREM VI. *If F and G are vector functions of the quantum variables $(p_1, p_2, p_3, q_1, q_2, q_3)$, then* (a)

$$F\cdot G - G\cdot F = \sum_{s=1} \frac{c^s}{s!}[H^{(s)}(G\cdot F) - H^{(s)}(F\cdot G)] \tag{25}$$

where

$$H^{(s)}(G\cdot F) = \sum_{\alpha+\beta+\gamma=s} \frac{s!}{\alpha!\beta!\gamma!} \frac{\partial^s G}{\partial q_1^\alpha \partial q_2^\beta \partial q_3^\gamma} \cdot \frac{\partial^s F}{\partial p_1^\alpha \partial p_2^\beta \partial p_3^\gamma},$$

and (b)

$$F\times G + G\times F = -\sum_{s=1} \frac{c^s}{s!}[H^{(s)}(G\times F) + H^{(s)}(F\times G)] \tag{26}$$

where

$$H^{(s)}(G\times F) = \sum_{\alpha+\beta+\gamma=s} \frac{s!}{\alpha!\beta!\gamma!} \frac{\partial^s G}{\partial q_1^\alpha \partial q_2^\beta \partial q_3^\gamma} \times \frac{\partial^s F}{\partial p_1^\alpha \partial p_2\, \partial p_3^\gamma}.$$

An alternate form for these expressions may be obtained as in the second part of Theorem V.

An interesting special case of formula (26) arises when $G=F$. Then

$$F\times F = -\sum_{s=1} \frac{c^s}{s!} H^{(s)}(F\times F).$$

UNIVERSITY OF IOWA,
IOWA CITY, IOWA

THE BASIC POWER SERIES OF INTERPOLATION*

BY
GEORGE RUTLEDGE

1. **Introduction.** The coefficients of the polynomial determined by the $2n+1$ points

$$(1) \qquad (-nh, y_{-n}), \cdots, (-h, y_{-1}), (0, y_0), (h, y_1), \cdots, (nh, y_n)$$

have been studied extensively by the author in a paper† to be referred to hereafter as (I).

There is no essential loss of generality in placing $h=1$ and $y_0=0$ in (1). Then the polynomial, $P^{(2n)}(x)$, has the form (I, (31), (22)):

$$(2) \quad \sum_{i=1}^{n} \sum_{k=1}^{n-i+1} \sum_{j=i+k-1}^{n} j\binom{j+i-k-1}{2i-2} \frac{A_{ij}^{(n)}}{2D^{(n)}} (\Delta_{-k}^{[2i-1]} + \Delta_{k}^{[2i-1]}) x^{2i-1}$$

$$+ \sum_{i=1}^{n} \sum_{j=i}^{n} \binom{j+i-1}{2i-1} \frac{A_{ij}^{(n)}}{2D^{(n)}} \Delta_0^{[2i]} x^{2i}$$

$$+ \sum_{i=1}^{n} \sum_{k=1}^{n-i} \sum_{j=i+k}^{n} \binom{j+i-k-1}{2i-1} \frac{A_{ij}^{(n)}}{2D^{(n)}} (\Delta_{-k}^{[2i]} + \Delta_{k}^{[2i]}) x^{2i}.$$

For fixed i, j the numbers

$$(3) \qquad \frac{A_{ij}^{(n)}}{2D^{(n)}} = (-1)^{i+j} \frac{\binom{2n}{n-j}}{\binom{2n}{n}} \frac{1}{j^2} \sum^{(j)} \frac{1}{r_1^2 \cdots r_{i-1}^2},$$

where the summation indicated is that of products $i-1$ at a time of squared reciprocals of the first n integers excepting j, are convergent functions of n, as n is increased indefinitely. However, for fixed i, k, $i>1$, the terms of the sums with respect to j in (2) are not bounded as n and j simultaneously increase. For this reason it is difficult to determine directly whether or not these sums converge.

---

* Presented to the Society, August 29, 1929; received by the editors in April, 1929.

† These Transactions, vol. 26 (1924), pp.113–123.

If the values (1), $h=1$, $y_0=0$, are given for the infinite set of all positive and negative integers, the polynomials (2) constitute a sequence which may or may not be convergent. The polynomial (2) is identical with the partial sum of a Stirling series, and with the partial sum of a Gauss series, on condition that the terms of these series are grouped in pairs involving differences of order $2i-1$ and $2i$. The conditions of convergence of the series of Stirling and Gauss have been the subject of much recent literature.* It is known, for example, that whenever the polynomials (2) constitute a convergent sequence for some one non-integral value of x, this sequence is convergent for every value of x to an integral function.†

In this paper the sums with respect to j in (2) will be transformed into sums the terms of which are independent of n, and are also of constant sign. The fact of convergence of the positive term sums which then express the coefficients of $(\Delta_{-k}^{[m]}+\Delta_k^{[m]})x^m$, and of $\Delta_0^{[2i]}x^{2i}$ in this new form of (2) constitutes the main result of the paper.

The transformed form (39) of (2) is a triple series the partial sums of which when taken as indicated are identical with partial sums of the series of Gauss and Stirling (with the understanding that in these series terms are grouped in pairs as stated above). When this triple series is absolutely convergent its sum taken in the order j, k, i is a power series representation of the integral function which is then defined by the series of Gauss and Stirling. It seems appropriate to refer to this power series as *the basic power series of interpolation.*

2. **Generating functions $P_{ik}^{(2n-1)}, P_{ik}^{(2n)}$.** If the Lagrange polynomial (2) is an odd function it may be written (I, (3))

$$\sum_{i=1}^{n}\sum_{j=1}^{n} j(y_j - y_{-j})\frac{A_{ij}^{(n)}}{2D^{(n)}}x^{2i-1}, \tag{4}$$

and if by reference to (2) we place

$$y_j = -y_{-j} = \binom{j+i-k-1}{2i-2}, \quad j = i+k-1, \cdots, n, \tag{5}$$
$$= 0, \quad j < i+k-1,$$

we obtain a polynomial for which the differences of order $2i-1$ are all zero with the exception

$$\Delta_k^{[2i-1]} = \Delta_{-k}^{[2i-1]} = 1. \tag{6}$$

* Cf. Nörlund, *Differenzenrechnung*, Berlin, 1924, pp. 208–222, for references to this literature.

† Nörlund, loc. cit., pp. 208–209.

We shall call this polynomial a *generating function*, and shall denote it by the symbol $P_{ik}^{(2n-1)}$. With the exception noted its differences of order $2i-1$ are all zero for all values of n.

Similarly, if the Lagrange polynomial (2) is an even function it may be written (I, (3))

$$\sum_{i=1}^{n}\sum_{j=1}^{n}(y_j + y_{-j})\frac{A_{ij}^{(n)}}{2D^{(n)}}x^{2i}, \tag{7}$$

and if by reference to (2) we place

$$y_j = y_{-j} = \binom{j+i-k-1}{2i-1}, \quad j = i+k, \cdots, n, \tag{8}$$

$$= 0, \quad j < i+k,$$

we obtain a polynomial for which the differences of order $2i$ are all zero with the exception

$$\Delta_k^{[2i]} = \Delta_{-k}^{[2i]} = 1, \quad k > 0; \tag{9}$$

or

$$\Delta_0^{[2i]} = 2, \quad k = 0. \tag{9a}$$

We shall also call this polynomial a *generating function*, and shall denote it by the symbol $P_{ik}^{(2n)}$. With the exception noted its differences of order $2i$ are all zero for all values of n.

It may be observed that the generating functions $P_{ik}^{(2n-1)}$, $P_{ik}^{(2n)}$ have properties *with respect to differences* of order $2i-1$, $2i$, respectively, wholly analogous to the properties *with respect to values* which characterize the polynomials which enter as coefficients into the formula of Lagrange in its classical form.

3. **Central differences of the generating functions.** Since we shall make essential use of the expansions of $P_{ik}^{(2n-1)}$ and $P_{ik}^{(2n)}$ in the form of Stirling's series, which depend on central differences, we shall now compute these central differences from the given functional values, (5) and (8).

We shall first of all illustrate the general method by an example. Consider the generating function $P_{21}^{(2n-1)}$. Writing the values of this function, for $x = -n, \cdots, n$, and the successive differences of these values in horizontal rows, we have the following:

$$\cdots -\binom{3}{2} \quad -\binom{2}{2} \quad 0 \quad 0 \quad 0 \quad \binom{2}{2} \quad \binom{3}{2} \cdots$$

$$\cdots \binom{2}{1} \quad \binom{1}{1} \quad 0 \quad 0 \quad \binom{1}{1} \quad \binom{2}{1} \cdots$$

$$\cdots \quad -1 \quad -1 \quad 0 \quad 1 \quad 1 \quad \cdots$$

$$\cdots \quad 0 \quad 1 \quad 1 \quad 0 \quad \cdots$$

$$\cdot \quad \cdot \quad \cdot \quad \cdot$$

$$\cdot \quad \cdot \quad \cdot$$

Using the notation (I, (38)) for mean central differences of odd order which is commonly used in writing Stirling's series, we have

$$(10) \qquad \mu\Delta_1'' = 1, \quad \mu\Delta_1^{V} = -1, \quad \mu\Delta_1^{VII} = 2, \quad \mu\Delta_1^{IX} = -5, \cdots.$$

We may, of course, determine these mean central differences directly from the values of the function. Thus for $\mu\Delta_1^{VII}$ we have

$$(11) \quad -\binom{7}{0}\left[-\binom{3}{2}\right]+\binom{7}{1}\left[-\binom{2}{2}\right]-\cdots$$
$$+\binom{7}{5}\binom{2}{2}-\binom{7}{6}\binom{3}{2}+\binom{7}{7}\binom{4}{2}.$$

Then by use of the identities*

$$(12) \qquad \sum_{s=0}^{m-q}(-1)^s\binom{m}{q+s}\binom{n+s}{n}=\binom{m-n-1}{m-q},$$
$$\sum_{s=0}^{q}(-1)^s\binom{m}{q-s}\binom{n+s}{n}=\binom{m-n-1}{q},$$
$$\binom{2n}{n-k}-\binom{2n}{n-k-1}=\frac{2k+1}{2n+1}\binom{2n+1}{n-k},$$

we obtain

$$(13) \qquad \mu\Delta_1^{VII} = \binom{4}{2}-\binom{4}{1}=\frac{1}{5}\binom{5}{2}.$$

The identities (12) suffice for the determination of the mean central differences and central differences for all the generating functions $P_{ik}^{(2n-1)}$, $P_{ik}^{(2n)}$, the manner of application being precisely that of the foregoing example. It will suffice, therefore, if we tabulate results, as follows:

* Netto, *Combinatorik*, p. 255.

CENTRAL DIFFERENCES OF $P_{ik}^{(2n-1)}$

(14)

		$\mu\Delta_1'$	$\mu\Delta_1'''$	$\mu\Delta_1^{V}$	$\mu\Delta_1^{VII}$
$i=1$:	$P_{11}^{(2n-1)}$	$\frac{1}{1}\binom{1}{0}$	$-\frac{1}{3}\binom{3}{1}$	$\frac{1}{5}\binom{5}{2}$	$-\frac{1}{7}\binom{7}{3}\cdots$
	$P_{12}^{(2n-1)}$		$\frac{3}{3}\binom{3}{0}$	$-\frac{3}{5}\binom{5}{1}$	$\frac{3}{7}\binom{7}{2}\cdots$
	$P_{13}^{(2n-1)}$			$\frac{5}{5}\binom{5}{0}$	$-\frac{5}{7}\binom{7}{1}\cdots$
			. . . . .		
			$\mu\Delta_1'''$	$\mu\Delta_1^{V}$	$\mu\Delta_1^{VII}$
$i=2$:	$P_{21}^{(2n-1)}$		$\frac{1}{1}\binom{1}{0}$	$-\frac{1}{3}\binom{3}{1}$	$\frac{1}{5}\binom{5}{2}\cdots$
	$P_{22}^{(2n-1)}$			$\frac{3}{3}\binom{3}{0}$	$-\frac{3}{5}\binom{5}{1}\cdots$
	$P_{23}^{(2n-1)}$				$\frac{5}{5}\binom{5}{0}\cdots$
				. . . . .	
				$\mu\Delta_1^{V}$	$\mu\Delta_1^{VII}$
$i=3$:	$P_{31}^{(2n-1)}$			$\frac{1}{1}\binom{1}{0}$	$-\frac{1}{3}\binom{3}{1}\cdots$
	$P_{32}^{(2n-1)}$				$\frac{3}{3}\binom{3}{0}\cdots$
				. . . . .	

CENTRAL DIFFERENCES OF $P_{ik}^{(2n)}$

(15)

		Δ_0''	Δ_0^{IV}	Δ_0^{VI}	Δ_0^{VIII}
$i=1$:	$P_{10}^{(2n)}$	$2\binom{0}{0}$	$-2\binom{2}{1}$	$2\binom{4}{2}$	$-2\binom{6}{3}\cdots$
	$P_{11}^{(2n)}$		$2\binom{2}{0}$	$-2\binom{4}{1}$	$2\binom{6}{2}\cdots$
	$P_{12}^{(2n)}$			$2\binom{4}{0}$	$-2\binom{6}{1}\cdots$
				. . . . .	

$i = 2$:

	Δ_0^{IV}	Δ_0^{VI}	Δ_0^{VIII}
$P_{20}^{(2n)}$	$2\binom{0}{0}$	$-2\binom{2}{1}$	$2\binom{4}{2}\cdots$
$P_{21}^{(2n)}$		$2\binom{2}{0}$	$-2\binom{4}{1}\cdots$
$P_{22}^{(2n)}$			$2\binom{4}{0}\cdots$

. . . . . .

$i = 3$:

	Δ_0^{VI}	Δ_0^{VIII}
$P_{30}^{(2n)}$	$2\binom{0}{0}$	$-2\binom{2}{1}\cdots$
$P_{31}^{(2n)}$		$2\binom{2}{0}\cdots$

. . . . . .

4. The Stirling series for $P_{ik}^{(2n-1)}$, $P_{ik}^{(2n)}$. Corresponding to the mean differences (14) we have the following Stirling series expansions:

$$(16)\quad P_{11}^{(2n-1)} = \frac{1}{1}\binom{1}{0}\frac{1}{1!}x - \frac{1}{3}\binom{3}{1}\frac{1}{3!}x(x^2-1) + \frac{1}{5}\binom{5}{2}\frac{1}{5!}x(x^2-1)(x^2-2^2) - \cdots,$$

$$P_{12}^{(2n-1)} = \frac{3}{3}\binom{3}{0}\frac{1}{3!}x(x^2-1) - \frac{3}{5}\binom{5}{1}\frac{1}{5!}x(x^2-1)(x^2-2^2) + \cdots,$$

$$P_{13}^{(2n-1)} = \frac{5}{5}\binom{5}{0}\frac{1}{5!}x(x^2-1)(x^2-2^2) - \cdots,$$

. . . . . ,

$$P_{21}^{(2n-1)} = \frac{1}{1}\binom{1}{0}\frac{1}{3!}x(x^2-1) - \frac{1}{3}\binom{3}{1}\frac{1}{5!}x(x^2-1)(x^2-2^2) + \cdots,$$

$$P_{22}^{(2n-1)} = \frac{3}{3}\binom{3}{0}\frac{1}{5!}x(x^2-1)(x^2-2^2) - \cdots,$$

. . . . . .

These expansions are then reducible to the following form:

$$(17)\quad P_{11}^{(2n-1)} = \frac{1}{1}\binom{1}{0}\frac{2}{\binom{2}{1}}\,\frac{x}{1^2} - \frac{1}{3}\binom{3}{1}\frac{4}{\binom{4}{2}}\,\frac{x(x^2-1)}{1^2 2^2} + \frac{1}{5}\binom{5}{2}\frac{6}{\binom{6}{3}}\,\frac{x(x^2-1)(x^2-2^2)}{1^2 2^2 3^2} + \cdots,$$

$$P_{12}^{(2n-1)} = \frac{3}{3}\binom{3}{0}\frac{4}{\binom{4}{2}}\,\frac{x(x^2-1)}{1^2 2^2} - \frac{3}{5}\binom{5}{1}\frac{6}{\binom{6}{3}}\,\frac{x(x^2-1)(x^2-2^2)}{1^2 2^2 3^2} + \cdots,$$

$$P_{13}^{(2n-1)} = \frac{5}{5}\binom{5}{0}\frac{6}{\binom{6}{3}}\,\frac{x(x^2-1)(x^2-2^2)}{1^2 2^2 3^2} - \cdots,$$

$$\cdots\cdots\cdots\cdots\cdots\cdots,$$

$$P_{21}^{(2n-1)} = \frac{1}{1}\binom{1}{0}\frac{4}{\binom{4}{2}}\,\frac{x(x^2-1)}{1^2 2^2} - \frac{1}{3}\binom{3}{1}\frac{6}{\binom{6}{3}}\,\frac{x(x^2-1)(x^2-2^2)}{1^2 2^2 3^2} + \cdots,$$

$$P_{22}^{(2n-1)} = \frac{3}{3}\binom{3}{0}\frac{6}{\binom{6}{3}}\,\frac{x(x^2-1)(x^2-2^2)}{1^2 2^2 3^2} - \cdots,$$

$$\cdots\cdots\cdots\cdots\cdots\cdots.$$

In like manner we have Stirling series expansions corresponding to the central differences (15), as follows:

$$(18)\quad P_{10}^{(2n)} = 2\binom{0}{0}\frac{1}{2!}x^2 - 2\binom{2}{1}\frac{1}{4!}x^2(x^2-1) + 2\binom{4}{2}\frac{1}{6!}x^2(x^2-1)(x^2-2^2) - \cdots,$$

$$P_{11}^{(2n)} = 2\binom{2}{0}\frac{1}{4!}x^2(x^2-1) - 2\binom{4}{1}\frac{1}{6!}x^2(x^2-1)(x^2-2^2) + \cdots,$$

$$P_{12}^{(2n)} = 2\binom{4}{0}\frac{1}{6!}x^2(x^2-1)(x^2-2^2) - \cdots,$$

$$\cdots\cdots\cdots\cdots,$$

$$P_{20}^{(2n)} = 2\binom{0}{0}\frac{1}{4!}x^2(x^2-1) - 2\binom{2}{1}\frac{1}{6!}x^2(x^2-1)(x^2-2^2) + \cdots,$$

$$P_{21}^{(2n)} = 2\binom{2}{0}\frac{1}{6!}x^2(x^2-1)(x^2-2^2) - \cdots,$$

$$\cdots\cdots\cdots\cdots.$$

The expansions (18) are reducible as follows:

$$(19)\quad P_{10}^{(2n)} = \binom{0}{0}\frac{2}{\binom{2}{1}}\frac{x^2}{1^2} - \binom{2}{1}\frac{2}{\binom{4}{2}}\frac{x^2(x^2-1)}{1^2 2^2} + \binom{4}{2}\frac{2}{\binom{6}{3}}\frac{x^2(x^2-1)(x^2-2^2)}{1^2 2^2 3^2} - \cdots,$$

$$P_{11}^{(2n)} = \binom{2}{0}\frac{2}{\binom{4}{2}}\frac{x^2(x^2-1)}{1^2 2^2} - \binom{4}{1}\frac{2}{\binom{6}{3}}\frac{x^2(x^2-1)(x^2-2^2)}{1^2 2^2 3^2} + \cdots,$$

$$P_{12}^{(2n)} = \binom{4}{0}\frac{2}{\binom{6}{3}}\frac{x^2(x^2-1)(x^2-2^2)}{1^2 2^2 3^2} - \cdots,$$

$$\cdots\cdots\cdots\cdots,$$

$$P_{20}^{(2n)} = \binom{0}{0}\frac{2}{\binom{4}{2}}\frac{x^2(x^2-1)}{1^2 2^2} - \binom{2}{1}\frac{2}{\binom{6}{3}}\frac{x^2(x^2-1)(x^2-2^2)}{1^2 2^2 3^2} + \cdots,$$

$$P_{21}^{(2n)} = \binom{2}{0}\frac{2}{\binom{6}{3}}\frac{x^2(x^2-1)(x^2-2^2)}{1^2 2^2 3^2} - \cdots,$$

$$\cdots\cdots\cdots\cdots.$$

From the definition of $P_{ik}^{(2n-1)}$ and by reference to (2), it follows that, for fixed i, k, the coefficient of x^{2i-1} in the expansion (17) of $P_{ik}^{(2n-1)}$ is equal to

$$2\sum_{j=i+k-1}^{n} j\binom{j+i-k-1}{2i-2}\frac{A_{ij}^{(n)}}{2D^{(n)}}. \tag{20}$$

Similarly, from the definition of $P_{ik}^{(2n)}$ and by reference to (2), it follows that, for fixed i, k, the coefficient of x^{2i} in the expansion (19) of $P_{ik}^{(2n)}$ is equal to

$$2\sum_{j=i+k}^{n}\binom{j+i-k-1}{2i-1}\frac{A_{ij}^{(n)}}{2D^{(n)}}. \tag{21}$$

For fixed i, k, $i>1$, the terms of (20) and (21) are not bounded as j and n simultaneously increase, but the equivalent expressions derived from (17) and (19) consist of terms *of constant sign* and *independent of n*. This we shall consider in the next section.

5. **The coefficients of x^{2i-1}, x^{2i} in $P_{ik}^{(2n-1)}$, $P_{ik}^{(2n)}$.** In this section we shall write down from (17) and (19) the coefficients which are equal to (20) and (21) respectively. We shall consider first the odd powers x^{2i-1}, then the even powers x^{2i}. For any odd power we have $k=1, 2, \cdots$, and for any even power we have $k=0, 1, \cdots$. The tabulation, arranged according to values of i and k, is derived directly in an evident manner from the expansions 17) and (19).

$P_{1k}^{(2n-1)}$ (*Coefficient of* x):

(22) $k=1$ $$\left\{\frac{1}{1}\binom{1}{0}\frac{2}{\binom{2}{1}}\frac{1}{1^2}+\frac{1}{3}\binom{3}{1}\frac{4}{\binom{4}{2}}\frac{1}{2^2}+\frac{1}{5}\binom{5}{2}\frac{6}{\binom{6}{3}}\frac{1}{3^2}+\cdots\right\},$$

$k=2$ $$-\left\{\frac{3}{3}\binom{3}{0}\frac{4}{\binom{4}{2}}\frac{1}{2^2}+\frac{3}{5}\binom{5}{1}\frac{6}{\binom{6}{3}}\frac{1}{3^2}+\cdots\right\},$$

. . . . . .

$P_{2k}^{(2n-1)}$ (*Coefficient of* x^3):

$$(23)\quad k=1\quad \left\{\frac{1}{1}\binom{1}{0}\frac{4}{\binom{4}{2}}\frac{1}{2^2}\sum_1^1\frac{1^*}{r_1^2}+\frac{1}{3}\binom{3}{1}\frac{6}{\binom{6}{3}}\frac{1}{3^2}\sum_1^2\frac{1}{r_1^2}+\cdots\right\},$$

$$k=2\quad -\left\{\frac{3}{3}\binom{3}{0}\frac{6}{\binom{6}{3}}\frac{1}{3^2}\sum_1^2\frac{1}{r_1^2}+\cdots\right\},$$

. . . . . .

$P_{3k}^{(2n-1)}$ (*Coefficient of* x^5):

$$(24)\quad k=1\quad \left\{\frac{1}{1}\binom{1}{0}\frac{6}{\binom{6}{3}}\frac{1}{3^2}\sum_1^2\frac{1^*}{r_1^2r_2^2}+\frac{1}{3}\binom{3}{1}\frac{8}{\binom{8}{4}}\frac{1}{4^2}\sum_1^3\frac{1}{r_1^2r_2^2}+\cdots\right\},$$

$$k=2\quad -\left\{\frac{3}{3}\binom{3}{0}\frac{8}{\binom{8}{4}}\frac{1}{4^2}\sum_1^3\frac{1}{r_1^2r_2^2}+\cdots\right\},$$

. . . . . .

$P_{4k}^{(2n-1)}$ (*Coefficient of* x^7):

$$(25)\quad k=1\quad \left\{\frac{1}{1}\binom{1}{0}\frac{8}{\binom{8}{4}}\frac{1}{4^2}\sum_1^3\frac{1^*}{r_1^2r_2^2r_3^2}+\frac{1}{3}\binom{3}{1}\frac{10}{\binom{10}{5}}\frac{1}{5^2}\sum_1^4\frac{1}{r_1^2r_2^2r_3^2}+\cdots\right\},$$

$$k=2\quad -\left\{\frac{3}{3}\binom{3}{0}\frac{10}{\binom{10}{5}}\frac{1}{5^2}\sum_1^4\frac{1}{r_1^2r_2^2r_3^2}+\cdots\right\},$$

. . . . . .

* Sums reducing to a single term are included to preserve symmetry.

$P_{1k}^{(2n)}$ (*Coefficient of* x^2):

$$(26)\quad k=0\quad \left\{\binom{0}{0}\frac{2}{\binom{2}{1}}\frac{1}{1^2} + \binom{2}{1}\frac{2}{\binom{4}{2}}\frac{1}{2^2} + \binom{4}{2}\frac{2}{\binom{6}{3}}\frac{1}{3^2} + \cdots\right\},$$

$$k=1\quad -\left\{\binom{2}{0}\frac{2}{\binom{4}{2}}\frac{1}{2^2} + \binom{4}{1}\frac{2}{\binom{6}{3}}\frac{1}{3^2} + \cdots\right\},$$

. . . . . .

$P_{2k}^{(2n)}$ (*Coefficient of* x^4):

$$(27)\quad k=0\quad \left\{\binom{0}{0}\frac{2}{\binom{4}{2}}\frac{1}{2^2}\sum_1^1\frac{1}{r_1^2} + \binom{2}{1}\frac{2}{\binom{6}{3}}\frac{1}{3^2}\sum_1^2\frac{1}{r_1^2} + \cdots\right\},$$

$$k=1\quad -\left\{\binom{2}{0}\frac{2}{\binom{6}{3}}\frac{1}{3^2}\sum_1^2\frac{1}{r_1^2} + \cdots\right\},$$

. . . . . .

$P_{3k}^{(2n)}$ (*Coefficient of* x^6):

$$(28)\quad k=0\quad \left\{\binom{0}{0}\frac{2}{\binom{6}{3}}\frac{1}{3^2}\sum_1^2\frac{1}{r_1^2r_2^2} + \binom{2}{1}\frac{2}{\binom{8}{4}}\frac{1}{4^2}\sum_1^3\frac{1}{r_1^2r_2^2} + \cdots\right\},$$

$$k=1\quad -\left\{\binom{2}{0}\frac{2}{\binom{8}{4}}\frac{1}{4^2}\sum_1^3\frac{1}{r_1^2r_2^2} + \cdots\right\},$$

. . . . . .

$P_{4k}^{(2n)}$ (*Coefficient of* x^8):

$$(29)\quad k=0\quad \left\{\binom{0}{0}\frac{2}{\binom{8}{4}}\frac{1}{4^2}\sum_1^3\frac{1}{r_1^2r_2^2r_3^2}+\binom{2}{1}\frac{2}{\binom{10}{5}}\frac{1}{5^2}\sum_1^4\frac{1}{r_1^2r_2^2r_3^2}+\cdots\right\},$$

$$k=1\quad -\left\{\binom{2}{0}\frac{2}{\binom{10}{5}}\frac{1}{5^2}\sum_1^4\frac{1}{r_1^2r_2^2r_3^2}+\cdots\right\},$$

. . . . . .

6. **Identities.** If we observe that

$$(30)\qquad \left|\frac{A_{ij}^{(j)}}{2D^{(j)}}\right|=\frac{1}{\binom{2j}{j}}\frac{1}{j^2}\sum_1^{j-1}\frac{1}{r_1^2\cdots r_{i-1}^2},$$

we may express the equality between (20), (21) and the corresponding sums appearing in (22) to (29) inclusive, as follows:

$$(31)\quad \sum_{j=i+k-1}^{n}\binom{j+i-k-1}{2i-2}j\frac{A_{ij}^{(n)}}{2D^{(n)}}=(-1)^{k+1}\sum_{j=i+k-1}^{n}\frac{2k-1}{2j-2i+1}\binom{2j-2i+1}{j-i-k+1}j\left|\frac{A_{ij}^{(j)}}{2D^{(j)}}\right|,$$

$$(32)\quad \sum_{j=i+k}^{n}\binom{j+i-k-1}{2i-1}\frac{A_{ij}^{(n)}}{2D^{(n)}}=(-1)^k\sum_{j=i+k}^{n}\binom{2j-2i}{j-i-k}\left|\frac{A_{ij}^{(j)}}{2D^{(j)}}\right|.$$

In (31) k has the values $1, 2, \cdots$, and in (32) the values $0, 1, 2, \cdots$. The significance of these identities lies first of all in the fact that the terms on the right are not functions of n and moreover are all of one sign. Of greater significance, however, is the fact that the positive term series on the right of these identities are convergent as n is increased indefinitely. This we shall now prove.

It may be seen at once that

$$\sum_{j=1}^{\infty} \binom{2j}{j} \left| \frac{A_{ij}^{(j)}}{2D^{(j)}} \right| = \sum_{j=1}^{\infty} \frac{1}{j^2} \sum_{1}^{j-1} \frac{1}{r_1^2 \cdots r_{i-1}^2} \tag{33}$$

is a particular arrangement of the products i at a time of the squared reciprocals of all positive integers. This sum is well known through its connection with

$$\frac{\sinh \pi x}{\pi x} = \prod_{1}^{\infty} \left(1 + \frac{x^2}{n^2}\right). \tag{34}$$

Therefore (33) is convergent, as follows:

$$\sum_{j=1}^{\infty} \binom{2j}{j} \left| \frac{A_{ij}^{(j)}}{2D^{(j)}} \right| = \frac{\pi^{2i}}{(2i+1)!}. \tag{35}$$

If we now place $j=i+k+s$, $s=0, 1, 2, \cdots$, we obtain for the ratio by which we may compare (31) with (33),

$$\frac{(2k+2s+1)\cdots(2k+s+1)}{(2k+2s+2i)\cdots(2k+s+2i)} \cdot \frac{(s+k+i)\cdots(s+2)}{(s+2k+2i-1)\cdots(s+k+i+1)} \frac{s+k+i}{2s+2k+1}(2k-1), \tag{36}$$

which is the value in fractional form of the ratio

$$\frac{2k-1}{2j-2i+1} \frac{j\binom{2j-2i+1}{j-i-k+1}}{\binom{2j}{j}}. \tag{37}$$

If in (36) we place $s=i+r$, $r=0, 1, 2, \cdots$, we see that for $j \geqq 2i+k$ the ratio (37) is less than $2k-1$. Thus (31) is convergent by comparison with (33).

If we compare (32) with (33), we see directly that for $j \geqq i+k$

$$\frac{\binom{2j-2i}{j-i-k}}{\binom{2j}{j}} < 1. \tag{38}$$

Thus the convergence of (32) is established.

7. **The power series of interpolation.** The identities (31), (32) serve to effect the transformation of (2) into the following form:

$$(39)\quad \sum_{i=1}^{n}\sum_{k=1}^{n-i+1}(-1)^{k+1}\sum_{j=i+k-1}^{n}\frac{2k-1}{2j-2i+1}\cdot\binom{2j-2i+1}{j-i-k+1}j\left|\frac{A_{ij}^{(j)}}{2D^{(j)}}\right|(\Delta_{-k}^{[2i-1]}+\Delta_{k}^{[2i-1]})x^{2i-1}$$

$$+\sum_{i=1}^{n}\sum_{j=i}^{n}\binom{2j-2i}{j-i}\left|\frac{A_{ij}^{(j)}}{2D^{(j)}}\right|\Delta_{0}^{[2i]}x^{2i}$$

$$+\sum_{i=1}^{n}\sum_{k=1}^{n-i}(-1)^{k}\sum_{j=i+k}^{n}\binom{2j-2i}{j-i-k}\left|\frac{A_{ij}^{(j)}}{2D^{(j)}}\right|(\Delta_{-k}^{[2i]}+\Delta_{k}^{[2i]})x^{2i}.$$

It has been proved in the preceding section that *the sums with respect to j in* (39) *are convergent positive-term series.* This theorem, together with the explicit form of (39), constitutes the main objective of this paper.

If the triple series (39) is absolutely convergent for a particular set of values (1), $h=1$, $y_0=0$, then the interpolation series of Stirling and Gauss are convergent (for this means only that the limit of the finite sum (39) exists as $n\to\infty$), and if (39) is summed in the order j, k, i it becomes the power series representation of the limit. It is possibly true conversely that (39) is absolutely convergent whenever the Gauss and Stirling series, with paired terms, are convergent. The series (39) has therefore a fundamental place in the theory of Gauss and Stirling series.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY,
CAMBRIDGE, MASS.

THE LIMIT POINTS OF A GROUP*

BY
LESTER R. FORD

1. **Introduction.** The present paper is concerned with properly discontinuous groups of linear transformations. Such a group can be so transformed that the point at infinity is carried into some fixed finite region by every transformation of the group other than the identical transformation. In what follows it will be understood, without express statement, that the groups referred to have this property.

The *isometric circle*† of a linear transformation

$$z' = T(z) = \frac{az + b}{cz + d}, \quad ad - bc = 1, \quad c \neq 0,$$

is the circle

$$|cz + d| = 1.$$

The isometric circles exist for all transformations of a group other than the identical transformation. For a given group these circles all lie in a finite region and have distinct centers; and they have the property that there is only a finite number with radius exceeding a given positive quantity.

By the *limit points* of a group is meant the cluster points of the centers of its isometric circles. It follows from the preceding remarks that there exist isometric circles belonging to the transformations of the group which lie wholly within any circle, however small, drawn with center at a limit point.

A group is called *elementary* if it has no limit points, or one limit point, or two limit points. If a group has more than two limit points, its limit points form a perfect set and so are non-denumerably infinite in number. It is with these latter, or non-elementary, groups that we shall be concerned.

2. **The fundamental theorem.** In the study of the limit points we shall make use of the properties of the isometric circles. The various results are based on the following theorem.

---

* Presented to the Society, December 28, 1927; received by the editors in February, 1929.

† For a summary of the properties of the isometric circle see Proceedings of the National Academy of Sciences, vol. 13 (1927), pp. 286–289. For a complete treatment see the author's *Automorphic Functions* (McGraw-Hill Book Company, 1929), §§11, 17 et seq.

THEOREM 1. *Let C and K be circles whose centers are the limit points P and Q respectively of a non-elementary group. Then there exists a transformation of the group whose isometric circle lies in C and the isometric circle of whose inverse lies in K.*

Let

$$I_1, I_2, I_3, \cdots$$

be a sequence of isometric circles belonging to the group and lying in C with centers

$$g_1, g_2, g_3, \cdots \rightarrow P ;$$

and let

$$J_1, J_2, J_3, \cdots$$

be a sequence lying in K with centers

$$h_1, h_2, h_3, \cdots \rightarrow Q.$$

Let T_n and S_n be the transformations of the group whose isometric circles are I_n and J_n respectively.

Paired with each isometric circle is the circle of equal radius belonging to the inverse transformation. Thus T_n carries I_n into I_n', the isometric circle of T_n^{-1}. Let

$$I_1', I_2', \cdots ; \qquad J_1', J_2', \cdots$$

represent the isometric circles of the inverses and let the centers of these be denoted by

$$g_1', g_2', \cdots ; \qquad h_1', h_2', \cdots .$$

Let G' be the set of cluster points of the set g_n', and H' the set of cluster points of the set h_n'. Each set has at least one cluster point, since it consists of an infinite set of points lying in a finite region.

We shall consider first the case $P \neq Q$. We may suppose (decreasing the circles, if necessary) that C and K are external to one another.

(1) Suppose that G' contains a point g' and H' a point h' where $g' \neq h'$. We can select subsequences of g_n' and h_n' approaching g' and h' respectively:

$$g_{n_1}', g_{n_2}', \cdots \rightarrow g' ; \quad h_{m_1}', h_{m_2}', \cdots \rightarrow h'.$$

We shall make repeated use of the following lemma, which has been proved elsewhere.*

* Proceedings of the National Academy of Sciences, vol. 13 (1927), p. 290; also *Automorphic Functions*, §24.

LEMMA. *Let I_t, I_t', I_s, I_s', and I_u be the isometric circles of T, T^{-1}, S, S^{-1}, and $U=TS$ respectively. If I_t and I_t' are exterior to one another then I_u is contained in I_s.*

Consider the transformation

$$T = S_{m_k}^{-1} T_{n_k}; \quad T^{-1} = T_{n_k}^{-1} S_{m_k}.$$

For k sufficiently large the isometric circles of $S_{m_k}^{-1}$ and $T_{n_k}^{-1}$, namely J_{m_k}' and I_{n_k}', are arbitrarily near h' and g' respectively and so are external. It follows from the lemma that the isometric circle of T lies in I_{n_k}, and hence in C. Similarly, applying the lemma to T^{-1}, the isometric circle of T^{-1} lies in J_{m_k} and in K. This proves the theorem for this case.

(2) Points g' and h' for the preceding treatment exist unless G' and H' each consists of a single point: $G'=H'=\alpha$.

If $\alpha=P$ or $\alpha=Q$ the theorem obviously holds for the transformation S_n or T_n for n sufficiently large. So we shall assume $\alpha\neq P, \alpha\neq Q$.

By taking n and m large enough we can make I_n as near to P, J_m as near to Q, and I_n' and J_m' as near to α as we wish. Hence we can choose n and m so that I_n and I_n' are exterior to one another and J_m and J_m' are exterior to one another. Now when the isometric circles of a transformation and its inverse are external circles the transformation is loxodromic or hyperbolic and each circle contains one of the fixed points of the transformation. Then I_n' contains a fixed point α_n of T_n and J_m' contains a fixed point β_m of S_m.

Suppose $\alpha_n\neq\beta_m$. Now α_n is a fixed point of T_n^q and β_m is a fixed point of S_m^q. If $q>1$ the isometric circle of T_n^q lies in I_n and that of S_m^q lies in J_m. Also the isometric circle of T_n^{-q} contains α_n and that of S_m^{-q} contains β_m. By taking q large enough we can make the radii of these circles less than $|\alpha_n-\beta_m|/4$. Then the isometric circles of T_n^{-q} and S_m^{-q} are external circles.

We now apply the lemma to the transformation $T=S_m^{-q}T_n^q$. Its isometric circle lies in that of T_n^q, and hence in I_n and in C. Similarly, the isometric circle of the inverse, $T^{-1}=T_n^{-q}S_m^q$, lies in that of S_m^q, and so in J_m and in K.

We now show that $\alpha_n=\beta_m$ is impossible for n and m large enough. The remaining fixed points of T_n and S_m are distinct since one lies near P and the other near Q. That $\alpha_n\neq\beta_m$ then follows directly from the following proposition.

If two hyperbolic or loxodromic transformations belonging to a group have one fixed point in common but the other fixed points distinct, then the group contains infinitesimal transformations.

Let the group be so transformed that the common fixed point is at infinity. Then the transformations have the forms

$$T: \quad z' - a = M(z - a), \qquad |M| \neq 1,$$
$$S: \quad z' - b = N(z - b), \qquad |N| \neq 1, \quad b \neq a;$$

and the powers of these transformations are

$$T^n(z) = M^n z + a(1 - M^n),$$
$$S^m(z) = N^m z + b(1 - N^m).$$

We now form the following product:

$$U = T^{-n}S^{-m}T^nS^m.$$

We find, on working this out, the translation

$$U = z + \omega,$$

where

$$\omega = (a - b)(1 - M^{-n})(1 - N^{-m}).$$

By suitable choices of m and n we can make M^{-n} and N^{-m} arbitrarily small and so make the period ω arbitrarily near $a-b$.

Given $\epsilon>0$, we form two such translations

$$U_1 = z + a - b + \epsilon_1, \quad U_2 = z + a - b + \epsilon_2,$$

where $|\epsilon_1| < \frac{1}{2}\epsilon$, $|\epsilon_2| < \frac{1}{2}\epsilon$, and $\epsilon_1 \neq \epsilon_2$. Then

$$U_1 U_2^{-1} = z + \epsilon_1 - \epsilon_2$$

is a translation of period $\epsilon_1 - \epsilon_2$, where $0 < |\epsilon_1 - \epsilon_2| < \epsilon$. Since ϵ may be chosen arbitrarily small, the group contains infinitesimal transformations. It is therefore not a discontinuous group.

There remains the case $P=Q$. We wish to prove that there exists a transformation T whose isometric circle and the isometric circle of whose inverse both lie in the circle C surrounding P.

Let C be made small enough that it does not contain all the limit points, and let P_1 and P_2 be distinct limit points outside C. About P_1 and P_2 construct circles C_1 and C_2 exterior to one another and to C.

Let T_1 be a transformation such that its isometric circle I_1 lies in C and that of its inverse, I_1', lies in C_1, and let T_2 be a transformation such that its isometric circle I_2 lies in C and that of its inverse, I_2', lies in C_2. We now apply the lemma to the transformation

$$T = T_2^{-1}T_1, \qquad T^{-1} = T_1^{-1}T_2.$$

Since I_2' and I_1' are exterior to one another the isometric circle of T lies in I_1 and hence in C. Similarly, the isometric circle of T^{-1} lies in I_2 and hence in C. This completes the proof of the theorem.

COROLLARY. *Let P and Q be limit points of a non-elementary group. Then there exists a transformation of the group, one of whose fixed points is in an arbitrarily small neighborhood of P and the other is in an arbitrarily small neighborhood of Q.*

This follows from the fact that the fixed points of a transformation lie within or on the isometric circles of the transformation and its inverse. In this corollary P and Q are not necessarily distinct.

We are now able to prove the following precise theorem embodying an important property of the limit points.

THEOREM 2. *Let C be a circle with center at a limit point P of a non-elementary group, and let Q be any point whatever in the plane. Then C contains an infinite number of distinct points congruent to Q.*

It suffices to prove that C contains one point congruent to Q and distinct from P; for we can then replace C by a circle C' with center P and small enough to exclude the first congruent point and thus get a second congruent point, and so step by step get an infinite number of distinct congruent points.

Let P_1 be a limit point lying in C and distinct from P and Q (here Q and P may be identical). Such a point exists, since P is a cluster point of limit points. Let P_2 be any other limit point distinct from P and Q. Let C_1 and C_2 be circles with centers P_1 and P_2 respectively and sufficiently small that P and Q are exterior to both circles and that C_1 lies in C. Let T be a transformation whose isometric circle I lies in C_2 and the isometric circle I' of whose inverse lies in C_1 (Theorem 1). T carries the exterior of I into the interior of I'. Hence Q is carried into a point Q' lying in I'. Then Q' lies in C and is distinct from P; and the theorem is established.

This theorem is no longer true if we drop the requirement that the group be non-elementary. Its failure occurs in the case of an elementary group if Q is a limit point.

3. **Non-loxodromic groups.** A properly discontinuous group is called Fuchsian if the transformations have a common fixed circle and each carries the interior of the circle into itself. It is well known that a Fuchsian group contains no loxodromic transformation. That a group without loxodromic transformations is not necessarily Fuchsian is shown by the doubly periodic group. This is, however, an elementary group. We shall prove the following

THEOREM 3. *A non-elementary properly discontinuous group which contains no loxodromic transformations is a Fuchsian group.*

We shall show first that all the limit points lie on a circle. Let A, B, C be distinct limit points and let K be the circle determined by them. We shall

show that any other limit point P lies on K. The necessary and sufficient condition for this is that the cross ratio

$$\lambda = \frac{(A - B)(C - P)}{(A - C)(B - P)}$$

be real.

The transformations

$$T(z) = (az + b)/(cz + d), \quad ad - bc = 1,$$
$$S(z) = (\alpha z + \beta)/(\gamma z + \delta), \quad \alpha\delta - \beta\gamma = 1,$$

are non-loxodromic if and only if $a+d$ and $\alpha+\delta$ are real. Let I_t, I_t', I_s, I_s' be the isometric circles of T, T^{-1}, S, S^{-1} respectively. The centers of these circles are $-d/c$, a/c, $-\delta/\gamma$, α/γ respectively. The first pair have radius $1/|c|$, the second pair $1/|\gamma|$.

According to Theorem 1 we can find transformations T and S belonging to the group so that a/c, $-d/c$, α/γ, $-\delta/\gamma$ are arbitrarily near A, C, B, P respectively and so that $1/|c|$ and $1/|\gamma|$ are arbitrarily small. Given $\epsilon>0$, we can choose T and S so that

$$\mu = \frac{\left(\frac{a}{c} - \frac{\alpha}{\gamma}\right)\left(-\frac{\alpha}{c} + \frac{\delta}{\gamma}\right)}{\left(\frac{a}{c} + \frac{d}{c}\right)\left(\frac{\alpha}{\gamma} + \frac{\delta}{\gamma}\right)} = \lambda + \epsilon'$$

where $|\epsilon'| < \frac{1}{2}\epsilon$, and also so that $|\epsilon''| < \frac{1}{2}\epsilon$, where

$$\epsilon'' = \frac{\frac{1}{c^2} + \frac{1}{\gamma^2}}{\left(\frac{a}{c} + \frac{d}{c}\right)\left(\frac{\alpha}{\gamma} + \frac{\delta}{\gamma}\right)}.$$

Now by hypothesis $S^{-1}T$ is non-loxodromic. We find readily that the condition for this is that

$$-\delta a + \beta c + \gamma b - \alpha d$$

be real. Hence

$$\nu = -\frac{-\delta a + \beta c + \gamma b - \alpha d}{(a + d)(\alpha + \delta)}$$

is real. On dividing by γc and using the relations $ad-bc=1$, $\alpha\delta-\beta\gamma=1$, we find

$$\nu = \frac{\left(\frac{a}{c} - \frac{\alpha}{\gamma}\right)\left(-\frac{d}{c} + \frac{\delta}{\gamma}\right) + \frac{1}{c^2} + \frac{1}{\gamma^2}}{\left(\frac{a}{c} + \frac{d}{c}\right)\left(\frac{\alpha}{\gamma} + \frac{\delta}{\gamma}\right)} = \mu + \epsilon''.$$

We have then

$$\lambda = \nu - \epsilon' - \epsilon'', \quad |\lambda - \nu| < \epsilon.$$

Since ν is real, the imaginary part of λ cannot exceed ϵ in absolute value, where ϵ is arbitrarily small. It follows that λ is real, and P lies on K.

Each transformation of the group carries the set of limit points into itself. Hence each transformation carries K into itself.

It remains to show that each transformation carries the interior of K into itself. Now every non-loxodromic transformation carries the interior of any of its fixed circles into itself with the exception of the elliptic transformation of period two ($a+d=0$) with its two fixed points on the fixed circle. We shall show that no transformations of this type occur.

Let T be such a transformation and let I be its isometric circle. Since T carries the interior of I into its exterior there are limit points both within and without I. Let P_1 and P_2 be limit points outside I; and let S be a transformation such that its isometric circle I_s lies near P_1 and that of its inverse I_s' is near P_2, and such that I, I_s, and I_s' are exterior to one another.

We consider the transformation

$$U = ST, \quad U^{-1} = T^{-1}S^{-1} = TS^{-1}.$$

Applying the lemma we have that the isometric circles I_u and I_u' lie respectively in I and I_s' and so are exterior to one another. Both S and U are by hypothesis non-loxodromic and are hence hyperbolic, since the isometric circles of the transformation and its inverse are exterior to one another. Both then carry the interior of K into itself. But this is impossible. Since T carries the interior of K into its exterior and S carries the exterior of K into the exterior, ST carries the interior into the exterior. This contradiction rules out the existence of the elliptic transformation T; and the theorem is established.

4. **Schottky sub-groups.** A Schottky group is a group generated by linear transformations $T_1, \cdots, T_n$, where T_k carries a circle Q_k into a circle Q_k' in such a way that the exterior of Q_k goes into the interior of Q_k' and where the $2n$ circles $Q_1, Q_1', \cdots, Q_n, Q_n'$ are exterior to one another. The limit points of the group lie in the $2n$ circles and form a discrete set of points.

THEOREM 4. *Given a non-elementary properly discontinuous group* Γ, *and a positive constant* ϵ. *Then there exists a Schottky sub-group* Γ' *of* Γ *such that every limit point of* Γ *is within a distance* ϵ *of a limit point of* Γ'.

It is understood here, as stated in the first paragraph of this paper, that the point at infinity is an ordinary point. Let the plane be divided into squares of side $\epsilon/3$ by rulings parallel to the real and imaginary axes. Then only a finite number N of these squares have limit points of Γ within them or on their boundaries. Let $P_1, P_1', \cdots, P_n, P_n'$ be distinct limit points of Γ so chosen that each of the N squares has one at least of the points within it or on its boundary. With $P_1, P_1', \cdots, P_n, P_n'$ as centers let circles C_1, $C_1', \cdots, C_n, C_n'$ respectively be drawn with radii sufficiently small that each circle is exterior to all the others, each radius in any case not exceeding $\epsilon/3$.

Let T_k be a transformation of Γ whose isometric circle I_k lies in C_k and the isometric circle I_k' of whose inverse lies in C_k' (Theorem 1). Then the circles $I_1, I_1', \cdots, I_n, I_n'$ are exterior to one another; and the transformations $T_1, \cdots, T_n$ generate a Schottky group.

Any limit point P of Γ is within a distance $2^{1/2}\,\epsilon/3$ of one of the points $P_1, \cdots, P_n'$ and each of these latter points is within a distance $\epsilon/3$ of limit points of Γ'. Hence P is within a distance ϵ of limit points of Γ'.

We note that the limit points of Γ' are also limit points of Γ.

RICE INSTITUTE,
HOUSTON, TEXAS

ON LINEAR UPPER SEMI-CONTINUOUS COLLECTIONS OF BOUNDED CONTINUA†

BY

WALLACE ALVIN WILSON

1. **Introduction and notation.** If $T=\{t\}$ denotes a closed set of points and with each point t there is associated a unique bounded continuum X(or X_t) in such a way that (a) $X_t \cdot X_{t'} = 0$ if $t \neq t'$, and (b) at each point $t=\tau$ of T the upper closed limit of X_t as $t \to \tau$ is a part of X_τ, we say that $X=f(t)$ is an upper semi-continuous function in T. The collection of continua $\{X\}$ is also known as an upper semi-continuous collection of continua. We say that $X=f(t)$ is a *minimal* upper semi-continuous function in T if there exists no upper semi-continuous function $Y=g(t)$ such that at every point t, $Y \subset X$ and at some point $Y \neq X$. Examples of this concept are given elsewhere.‡

The following notation will be convenient. If $X=f(t)$ in T and $M=\sum[X]$, we write $M=F(T)$. If T is a bounded continuum and $f(t)$ is upper semi-continuous, it is obvious that M is a bounded continuum; in this case we say that X is an *element* of M. If T is a simple arc ab, $M=F(ab)$ will be called a *generalized arc*, or simply an arc if no confusion is caused. This may be denoted by $X_a X_b$ and the elements X_a and X_b will be called the ends. Likewise, $M-(X_a+X_b)$ is called a (*generalized*) *open arc* and denoted by $X_a^* X_b^*$. The meaning of $X_a X_b^*$ and $X_a^* X_b$ is apparent. The terms "upper limit" and "limit" of a system of continua are used in the sense of the closed limits of Hausdorff (*Grundzüge der Mengenlehre*, p. 236) as extended by L. S. Hill (*Properties of certain aggregate functions*, American Journal of Mathematics, vol. 49, pp. 420–421). If K is the upper limit of X_t as $t \to \tau$, we write $K = \lim\sup_{t \to \tau} X_t$; if the domain of definition T is a linear set and we restrict t to points at the right or left of τ, we write $R \lim\sup_{t \to \tau} X_t$ or $L \lim\sup_{t \to \tau} X_t$, respectively.

In the article referred to above various properties of upper semi-continuous functions were derived and in particular the following theorem was proved:

---

† Presented to the Society, February 23, 1929; received by the editor of the Bulletin in January, 1929, accepted for publication in the Bulletin, and subsequently transferred to the Transactions.

‡ W. A. Wilson, *Some properties of upper semi-continuous collections of bounded continua*, Bulletin of the American Mathematical Society, vol. 34, pp. 599–606.

Let the aggregate $\{t\}$ be a simple arc ab, let $X=f(t)$ be a minimal upper semi-continuous function defined over ab, let $M=\sum[X]$ lie in a plane, and let no element of M separate X_a from X_b. Then the continuum M is irreducible between X_a and X_b.

The last two hypotheses are needed in the particular proof of the theorem that is given, but there is no evidence that they are essential for the validity of the conclusion. It is the purpose of this article to discuss this particular question and to give some properties of generalized arcs which lie in a plane and have one or more elements separating† the ends.

2. THEOREM. *Let $M=F(ab)$ be a generalized bounded arc in a plane Z. Let t_1, τ, and t_2 be points of ab, let $t_1<\tau<t_2$, and let X_τ separate $X_1=f(t_1)$ from $X_2=f(t_2)$. Then, if $t'<\tau<t''$, X_τ separates $X'=f(t')$ from $X''=f(t'')$.*

Let R and S be the components of $Z-X_\tau$ containing X_1 and X_2, respectively. Since $X_aX_\tau^*$ is connected and contains X_1 it lies in R. For a like reason $X_\tau^*X_b$ lies in S. But $X'\subset X_a X_\tau^*$ and $X''\subset X_\tau^*X_b$. Hence the theorem is proved.

Remarks. It should be noted that, if some sub-continuum K of M separates X_a from X_b, at least one element X has this property. (See reference under §1, p. 601.) As will be seen later, however, X need not be a part of K, nor K a part of X.

3. THEOREM. *Let $M=F(ab)$ be a generalized bounded arc in a plane Z. Let $X_1=f(t_1)$ and $X_2=f(t_2)$ separate X_a from X_b and $t_1<t_2$. Then X_1 separates X_a from X_2, and X_2 separates X_1 from X_b.*

Let the components R_1 and S_1 of $Z-X_1$ contain X_a and X_b, respectively. Then $R_1\supset X_aX_1^*$ and $S_1\supset X_1^*X_b$. Let R_2 and S_2 be likewise defined for X_2. Then $R_2\supset X_aX_2^*$ and $S_2\supset X_2^*X_b$. Since $t_1<t_2$, $X_1\subset X_aX_2^*\subset R_2$ and $X_2\subset X_1^*X_b\subset S_1$. This proves the theorem.

As a consequence of this theorem, it follows that, if the elements of M which separate X_a and X_b are arranged in the order of the corresponding values of t, each such element may be said to lie within all those which follow it and to lie without all those which precede it, or vice versa.

4. THEOREM. *Let $M=F(ab)$ be a generalized bounded arc in a plane. Let t' be any point of ab for which $X'=f(t')$ separates X_a from X_b. Then the set of points $\{t'\}$ together with a and b form a closed set.*

Let this set be T' and let τ be a limiting point of T' different from a

† Throughout this article the term "separate" has reference to the plane in which the aggregates referred to lie.

and b. Then there is a sequence $\{t_n'\}$ of points of $T'-(a+b)$ converging to τ. Let $X_n'=f(t_n')$. Since $f(t)$ is upper semi-continuous, $X_\tau \supset \lim\sup_{n\to\infty} X_n'$. By hypothesis each X_n' separates X_a from X_b. Since $(X_a+X_b)\cdot X_\tau=0$, $\lim\sup_{n\to\infty} X_n'$ separates X_a from X_b.† Hence τ lies in T'. As $a+b \subset T'$, this shows T' is closed.

5. THEOREM. *Let $X_t=f(t)$ be a minimal upper semi-continuous function defined over the interval ab. Let $M=F(ab)$ be bounded and lie in a plane. Let the set of points $T''=\{t'\}$ for which $X'=f(t')$ separates X_a from X_b be void or totally disconnected. Then M is irreducible between X_a and X_b.*

The case that T'' is void is merely the theorem cited in §1. In the alternative case set $T'=T''+a+b$. Then T' is a nowhere dense closed set and $ab-T'$ is a finite or enumerable set of open intervals whose end points are points of T'.

Let K be a sub-continuum of M irreducible between X_a and X_b. Let c^*d^* be any one of the open intervals whose sum is $ab-T'$. Set $g(t)=f(t)$ for $c<t<d$; $g(c)=R\lim\sup_{t\to c} f(t)$; and $g(d)=L\lim\sup_{t\to d} f(t)$. Then $g(t)$ is a minimal upper semi-continuous function in cd and it is clear that $G(cd)=\overline{X_c^*X_d^*}$. But for no t in c^*d^* does $g(t)$ separate $g(c)$ from $g(d)$ by the hypothesis regarding cd and §2. Then the theorem quoted in §1 shows that $\overline{X_c^*X_d^*}$ is irreducible between $g(c)$ and $g(d)$ and therefore between X_aX_c and X_dX_b. But K contains a sub-continuum irreducible between these arcs and $K \subset M$. Hence $K \supset \overline{X_c^*X_d^*}$.

Now set $Y_t=K\cdot X_t=k(t)$. We have just seen that for any point t in $ab-T'$, $Y_t=X_t$. Let t' be any point of T' except a or b. Since T' is nowhere dense, there are two univariant sequences of points $\{s_n\}$ and $\{t_n\}$ lying in $ab-T'$ and converging to t' such that $s_n<t'<t_n$; let M_n denote the arc of M whose ends are $f(s_n)$ and $f(t_n)$. Since $f(s_n)=k(s_n)$ and $f(t_n)=k(t_n)$, it is evident that $K\cdot M_n$ is a continuum. Obviously $X_{t'}=\prod_1^\infty[M_n]$. Then $Y_{t'}=K\cdot X_{t'}=\prod_1^\infty[K\cdot M_n]$, and is therefore a continuum. Similar reasoning shows that Y_a and Y_b are continua.

We now show that $Y_t=k(t)$ is upper semi-continuous. This needs a proof only for points t' of T', since $k(t)=f(t)$ in $ab-T'$. As $Y_t \subset K$ and $Y_t \subset X_t$, $\lim\sup_{t\to t'} Y_t \subset K\cdot\lim\sup_{t\to t'} X_t \subset K\cdot X_{t'}$. But $Y_{t'}=K\cdot X_{t'}$ by definition. Hence $\lim\sup_{t\to t'} Y_t \subset Y_{t'}$ everywhere.

Since $f(t)$ is a minimal upper semi-continuous function and $k(t)\subset f(t)$ everywhere, it follows that for every t, $Y_t=k(t)=f(t)=X_t$. Hence $K=M$ and the theorem is proved.

† For, if A and B are bounded continua; $\{X_n\}$ is a sequence of bounded continua, each of which separates A from B; $X=\lim\sup_{n\to\infty} X_n$; and $X\cdot(A+B)=0$; then X separates A from B.

Examples to which the theorem is applicable are to be found in the literature. Here is another. Let $a=0$ and $b=1$, and $T'=\{t'\}$ be any nowhere dense closed set in ab. For each t' in T' let $f(t')$ be a circumference whose center is X_0 and whose radius is t'; let $f(0)=X_0$ and $f(1)$ be a circumference with center X_0 and radius 1. Let t_1' and t_2' be any two points of T' which are the end points of one of the open intervals whose sum is $ab-T'$. In the ring bounded by $f(t_1')$ and $f(t_2')$ place a spiral approaching the bounding circumferences asymptotically and for any t between t_1' and t_2' meeting the circumference of radius t and center X_0 in a single point, which we take as $f(t)$. Do this for each open interval of $ab-T'$. Then $f(t)$ is a minimal upper semi-continuous function and $M=F(01)$ is irreducible between X_0 and X_1.

Let us now turn to the hypothesis that T' is totally disconnected. It is clear that, if $f(t)$ is merely upper semi-continuous in ab, there may be a sub-interval cd such that for every t in cd, X_t separates X_a from X_b. A set of concentric circumferences is an example, but in this case $f(t)$ would not be a *minimal* function. If it is impossible for $f(t)$ to have this quality and for T' at the same time to contain an interval, the hypothesis is redundant; on the other hand, the conclusion of the theorem may be true even if the hypothesis is not satisfied. Two examples bearing on these points will be given.

6. **Example I.**† Let K_i and L_i denote two simple closed plane curves having one common point x_i and let K_i-x_i lie within L_i. The union of K_i+L_i and the complementary domain whose frontier is this continuum we call a *crescent* and we denote it by C_i. The point x_i will be called a *cut point*.

Let R be a circular ring bounded by two concentric circumferences K and L, K being the smaller. It is easy to show that, if n is any integer, we can construct n crescents $C_1, C_2, \cdots, C_n$ in R which satisfy the following requirements. No two of the sets C_i have common points. $R-\sum[C_i]$ consists of $n+1$ rings (not necessarily circular) $R_0, R_1, \cdots, R_n$ whose frontiers are $K+K_1, L_1+K_2, \cdots, L_{n-1}+K_n, L_n+L$, respectively, and each of these rings has in common with every radius from the common center E of K and L a segment of uniform length w, which may be called the width of the ring. If E is taken for the pole of a system of polar coördinates, the vectorial angle θ of each cut point x_i is $2\pi i/n$. Finally this construction can be repeated in each of the rings $\{R_i\}$.

Set up a system of polar coördinates (r,θ), where θ is measured in terms

† This is a modification of a continuum previously described by the author: *A curious irreducible continuum*, Bulletin of the American Mathematical Society, vol. 32, pp. 679–681.

of 2π as a unit. Let $0 \leqq t \leqq 1$. Let $X_0 = f(0)$ and $X_1 = f(1)$ be the circumferences $r = 1$ and $r = 2$, respectively, and let R_0 be the ring bounded by these. As described in the previous paragraph, construct in R_0 two crescents C_1 and C_2, whose frontiers are $X_{0,1} = K_1 + L_1$ and $X_{0,2} = K_2 + L_2$ and which divide R_0 into three rings $R_{0,0}$, $R_{0,1}$, and $R_{0,2}$, each of which has the width ϵ_1. Set $X_{0,1} = f(1/3)$ and $X_{0,2} = f(2/3)$ and let the cut points of C_1 and C_2 have θ equal to 1/2 and 1, respectively.

In $R_{0,0}$ construct four crescents $C_{0,1}$, $C_{0,2}$, $C_{0,3}$, and $C_{0,4}$, which divide $R_{0,0}$ into five rings $R_{0,0,0}$, $R_{0,0,1}$, $R_{0,0,2}$, $R_{0,0,3}$, $R_{0,0,4}$, each of width ϵ_2, and whose cut points have θ equal to 1/4, 1/2, 3/4, and 1, respectively. Let their frontiers be $X_{0,0,1} = f(1/15)$, $X_{0,0,2} = f(2/15)$, $X_{0,0,3} = f(3/15)$, and $X_{0,0,4} = f(4/15)$. In the same way treat $R_{0,1}$ and $R_{0,2}$, getting $f(6/15)$, $f(7/15)$, $f(8/15)$, $f(9/15)$, $f(11/15)$, $f(12/15)$, $f(13/15)$, and $f(14/15)$.

In each of the 15 rings thus formed construct 8 crescents whose frontiers are $f(1/135)$, $f(2/135)$, $\cdots$, $f(8/135)$; $f(10/135)$, $\cdots$, $f(17/135)$; $\cdots$, $f(134/135)$. In each of these 15 rings let the cut points have θ equal to 1/8, 1/4, 3/8, 1/2, 5/8, 3/4, 7/8, and 1.

Let this process be continued indefinitely; it is clear that the sequence $\epsilon_1, \epsilon_2, \cdots$ converges to zero. This defines $X = f(t)$ for any rational $t < 1$ of the form p/q, where p is relatively prime to q and q is one of the integers $3, 3 \cdot 5, 3 \cdot 5 \cdot 9, 3 \cdot 5 \cdot 9 \cdot 17$, etc. Any other t is the divisor of a decreasing sequence of intervals $d_1, d_2, \cdots$, of lengths 1/3, 1/15, 1/135, etc., and to each of these corresponds a definite ring of width ϵ_i lying between the crescents corresponding to the end points of the interval. Let the divisor of such a sequence of rings be $f(t)$; it is obviously a simple closed curve.

The construction insures that $f(t)$ is upper semi-continuous. It is obvious that any continuum P joining X_0 to X_1 contains all the cut points. As these are everywhere dense with respect to θ *in each ring*, they are everywhere dense in $M = \sum [X]$. Hence $P = M$ and M is irreducible between X_0 and X_1. Hence also $f(t)$ is a minimal function.

Thus we have an example where $f(t)$ is a minimal upper semi-continuous function defined over an interval ab and for every t between a and b the corresponding element separates $f(a)$ from $f(b)$. That is, the hypothesis in the theorem of §5 that the set of points T' be totally disconnected is not redundant and the conclusion of this theorem may be valid with this hypothesis omitted. We now proceed to construct an example which shows that this last situation does not always happen.

7. **Example II.** As a preliminary we state several facts regarding bounded continua which are either well known or easily demonstrated from known theorems.

(a) For the sake of brevity later a bounded plane continuum F which has precisely two complementary domains R and S and is the frontier of both of them will be called a continuum of type α. For such a continuum and for any $\epsilon>0$, there exist simple polygons P and P' in R and S, respectively, which have these properties: $R+F$ lies in the interior of P'; $P+P' \subset V_\epsilon(F)$, $F \subset V_\epsilon(P)$, $F \subset V_\epsilon(P')$; the rings between P and F and between F and P' contain no circle of diameter ϵ; and the ring between P and P' contains no circle of diameter 2ϵ. (The notation $V_\epsilon(F)$ means the set of all points whose distances from F are less than ϵ.)

(b) It is possible to construct an indecomposable continuum K of type α such that there is no continuum C joining a point of R to one of S such that $C \cdot K$ is a proper sub-continuum of K. In other words $C \cdot K$ is always disconnected or identical with K. Such continua will be referred to later as of *type* β.†

(c) Let $\{F_i\}$ and $\{G_i\}$ be two sequences of continua of type α having the following properties. If $i<i'$, let F_i lie in the interior of $F_{i'}$, i.e., in the bounded complementary domain. Let every F_i be in the interior of every G_i. If $i<i'$, let $G_{i'}$ lie in the interior of G_i. For every $\epsilon>0$ let there be an i_0 such that for every $i>i_0$, $F_i \subset V_\epsilon(G_i)$, $G_i \subset V_\epsilon(F_i)$, and the ring H_i whose frontier is F_i+G_i contains no circle of diameter ϵ. Then, if $H=\prod_1^\infty [H_i]$, H is a continuum of type α separating every F_i from every G_i and $H=\lim_{i\to\infty} F_i=\lim_{i\to\infty} G_i$.

(d) Let P and P' be simple polygons, let P be in the interior of P', and let R be the ring between them. Then, for every $\epsilon>0$, there is an integer n, such that there are n polygons $\{P_i\}$ dividing R into $n+1$ rings, such that no ring contains a circle of diameter ϵ, each ring is contained in an ϵ-vicinity of each of the polygons which bound it, and, if we set $P_0=P$ and $P_{n+1}=P'$, every $P_i \subset V_\epsilon(P_{i+1})$ and every $P_{i+1} \subset V_\epsilon(P_i)$, $i=0, 1, 2, \cdots, n$.

We are now ready to proceed with our construction. Let K be a continuum of type β, let P' and P'' be simple polygons, let P' be in the interior of K and K in the interior of P'', and let Γ be the ring between P' and P''.

Let C be a circumference of center E and radius 1. Let $0<\epsilon_1<1/4$. Let $C_0=E$ and $C_1, C_2, \cdots, C_{n_1}=C$ be a set of concentric circumferences,

† It is natural to infer that any indecomposable continuum of type α is necessarily of type β. That this is not true may be seen from the following modification of the continua of Wada, the idea for which is due to a suggestion by H. M. Gehman. Let C' and C'' be two circumferences, C' being within C'', and let L be a segment joining C' and C''. Let the simply connected region whose frontier is $C'+C''+L$ be the "island" of K. Yoneyama (Tôhoku Mathematical Journal, vol. 12, pp. 60–62), the interior of C' be the "lake of fresh water," and the exterior of C'' be the "sea." If "canals" are constructed as described by Yoneyama, the resulting indecomposable continuum is of type α, but not of type β. For the constituent containing L is accessible from both of the complementary domains.

each within the following and having a distance from the two adjacent ones less than ϵ_1. Put Γ in a homeomorphism with each of the n_1-1 circular rings thus determined and let K_i, $i=1, 2, \cdots, n_1-1$, be the image of K. Set $K_0=E$ and $K_{n_1}=C$. For each $i<n_1$, $K_i \subset V_{\eta_1}(K_{i+1})$ and $K_{i+1} \subset V_{\eta_1}(K_i)$, where $\eta_1=2\epsilon_1$. Moreover, the point $K_0=E$ and the continua $K_1, K_2, \cdots, K_{n_1-1}$ divide the interior of C into n_1 rings $R_0, R_1, R_2, \cdots, R_{n_1-1}$ no one of which contains a circle of diameter η_1. At the same time divide the interval $0 \leqq t \leqq 1$ into n_1 equal intervals $D_0, D_1, D_2, \cdots, D_{n_1-1}$ by the points $c_1, c_2, \cdots, c_{n_1-1}$ and set $c_0=0$ and $c_{n_1}=1$. For each point $t=c_i$, set $f(c_i)=K_i$. We note that this also establishes a correspondence between the intervals $\{D_i\}$ and the rings $\{R_i\}$.

For some positive $\epsilon_2<\frac{1}{2}\epsilon_1$ and some integer n_2 there can be constructed, in each ring R_i, n_2 simple polygons dividing it into n_2+1 rings in such a manner that, if all the n_1n_2 polygons and the continua $\{K_i\}$ are taken in order starting from $K_0=E$, each polygon or continuum is contained in an ϵ_2-vicinity of the one preceding and the one following and no ring contains a circle of diameter ϵ_2. This is a consequence of (a) and (d), above. Each ring R_i contains n_2-1 rings bounded by polygons. Put Γ in a homeomorphism with each of these and let the images of K be $\{K_{i,j}\}$, $j=1, 2, \cdots, n_2-1$. Set $K_{i,0}=K_i$ and $K_{i,n_2}=K_{i+1}=K_{i+1,0}$. If $\eta_2=2\epsilon_2$, each $K_{i,j}$ is in an η_2-vicinity of $K_{i,j+1}$ and of $K_{i,j-1}$. Moreover, the n_1n_2 continua $\{K_{i,j}\}$, $i=0, 1, 2, \cdots, n_1-1$, $j=0, 1, 2, \cdots, n_2-1$ divide the interior of C into n_1n_2 rings $R_{0,0}, R_{0,1}, \cdots, R_{n_1-1,n_2-1}$ no one of which contains a circle of diameter η_2, and the frontier of each $R_{i,j}$ is $K_{i,j}+K_{i,j+1}$. At the same time divide each of the intervals D_i into n_2 equal intervals $\{D_{i,j}\}$, $j=0, 1, 2, \cdots, n_2-1$, by the points $c_{i,1}, c_{i,2}, \cdots, c_{i,n_2-1}$. For each point $c_{i,j}$ set $f(c_{i,j})=K_{i,j}$, $i=0, 1, 2, \cdots, n_1-1$, $j=0, 1, 2, \cdots, n_2-1$. This also establishes a correspondence between the intervals $\{D_{i,j}\}$ and the rings $\{R_{i,j}\}$.

Now take $\epsilon_3<\frac{1}{2}\epsilon_2$ and continue this process indefinitely. If t is one of the division points $c_{i,j,k}$, $c_{i,j,k,l}$, etc. set $f(t)$ equal to the corresponding $K_{i,j,k}$, $K_{i,j,k,l}$, etc. Any other point t is the divisor of a decreasing sequence of intervals D_i, $D_{i,j}$, $D_{i,j,k}$, etc., to which corresponds a decreasing sequence of rings R_i, $R_{i,j}$, $R_{i,j,k}$, etc., whose divisor is a continuum. This continuum we set equal to $f(t)$. It is clear from the construction that, if $M=\sum[f(t)]=F(01)$ and C' denotes the sum of C and its interior, $M=C'$.

Let us now consider the nature of $f(t)$. Since $\epsilon_n \to 0$, it follows from the construction and (c) above that $f(t)$ is not only upper semi-continuous, but continuous for each t not a division point c_i, $c_{i,j}$, $c_{i,j,k}$, etc. This is also true at the division points, for any such point, as $c_{i,j,k}$, can be regarded as the divisor of the sequence of intervals c_ic_{i+1}, $c_{i,j}c_{i,j+1}, c_{i,j,k-1}c_{i,j,k+1}$, $c_{i,j,k-1,n_4-1}$

$\cdot c_{i,j,k,1}$, etc., and the sequence of rings bounded by the pairs of continua corresponding to the end points of these intervals with their frontiers satisfy the hypotheses of (c). It also follows from (c) that, if t is different from 0 and 1, $f(t)$ separates $f(0)$ from $f(1)$.

Let $g(t) \subset f(t)$ and be upper semi-continuous in (01) and let $N = \sum [g(t)]$. The sets of continua K_i, $K_{i,j}$, $K_{i,j,k}$, etc. were constructed so as to be everywhere dense in M. Hence, if $M \neq N$, there is some K containing points not in N. Let this K be $f(\tau)$; then $g(\tau) \neq f(\tau)$. This gives us a continuum N joining the interior of $f(\tau)$ to the exterior and having in common with $f(\tau)$ a proper sub-continuum $g(\tau)$ of $f(\tau)$. This is impossible by (b) above since each K is of type β. Therefore $M = N$; i.e., $g(t) = f(t)$ at every point and $f(t)$ is a *minimal* upper semi-continuous function. As M fills a part of the plane, it is not an irreducible continuum.

8. **Remarks.** As already stated, the construction of Example II shows that, if $f(t)$ is a minimal upper semi-continuous function defined over an interval ab, $M = F(ab)$ need not be irreducible between X_a and X_b, even though M is a plane continuum. The example can be easily modified to give a similar construction in space. Whether such an example can be constructed in space without the use of cyclic continua is an open question. An example of a continuous decomposition of an irreducible continuum into elements none of which are points, communicated to the author by Dr. Bronislaw Knaster, makes it seem probable that a decomposition similar to that in Example II can be arrived at without the use of indecomposable continua. Finally, this example justifies the last sentence in §2.

YALE UNIVERSITY,
NEW HAVEN, CONN.

THE ASYMPTOTIC LOCATION OF THE ROOTS OF A CERTAIN TRANSCENDENTAL EQUATION*

BY

RUDOLPH E. LANGER

The determination of the characteristic values in many boundary problems associated with differential or integral equations depends upon the location of the roots of an equation of the form

$$|\phi_{ij}(\lambda) + \psi_{ij}(\lambda)e^{\lambda A_j}| = 0. \tag{1}$$

The symbol $|\alpha_{ij}|$ is used here to designate the determinant of order n with general element α_{ij}. In expanded form the equation (1) may be written

$$\sum \omega_l(\lambda)e^{\lambda B_l} = 0, \tag{2}$$

where each B_l is a combination of the values A_j, and ω_l is expressible in terms of the coefficients ϕ_{ij}, ψ_{ij}.

The following are the hypotheses to be made concerning equation (1):

(i) the exponents A_j are complex constants;

(ii) the functions $\phi_{ij}(\lambda)$, $\psi_{ij}(\lambda)$ are such that the coefficients $\omega_l(\lambda)$ of the equation in the expanded form (2) are analytic for $|\lambda|$ sufficiently large, and are of the form

$$\omega_l(\lambda) = \lambda^{\nu_l}[a_l], \tag{3}$$

where ν_l is real and a_l is a constant not zero. If ν is not an integer, λ^ν is taken to mean $\exp(\nu \log \lambda)$, the principal value of the logarithm to be chosen.

The symbol $[a_l]$ in (3) designates a quantity $a_l + \epsilon(\lambda)$, where $\epsilon(\lambda)$ is used as a generic designation for a function which is analytic for $|\lambda|$ sufficiently large and which approaches zero uniformly as $|\lambda| \to \infty$.

The equation of the type (1) either in form (1) or form (2) has been extensively studied under hypotheses of varying degrees of restrictiveness, by Tamarkin,† Wilder,‡ and Pólya and Schwengeler.§ Of these, the two last

---

* Presented to the Society, March 30, 1929. Received by the editor of the Bulletin in March 1929, accepted for publication in the Bulletin, and subsequently transferred to these Transactions.

† (1) *Some General Problems of the Theory of Ordinary Linear Differential Equations* etc., Petrograd, 1917 (in Russian), also (2), Mathematische Zeitschrift, vol. 27 (1927), p. 24.

‡ These Transactions, vol. 18 (1917), p. 415.

§ Pólya, Sitzungsberichte der Bayerischen Akademie, München, 1920, p. 285.

Schwengeler, E., *Geometrisches über die Verteilung der Nullstellen* etc., Zürich, 1925.

named focus the attention in large measure upon the geometric determination of the configuration of regions in which the roots of the equation are located. They consider the case in which the coefficients are polynomials; a restriction which would appear from the following discussion to be unessential. The present note is intended as a brief analytic rather than geometric discussion. The final result is more precise than that obtained by Schwengeler.* The deductions are based directly on the work of the other authors cited above. This work thus appears, indeed, to include in large measure the analytic deductions on the location of the roots made subsequently by Schwengeler.

The constants A_j are by hypothesis complex. We designate by $\mu_1, \mu_2, \cdots, \mu_r$ the distinct values of the set $(\arg A_j)$ arranged in the order

$$0 \leqq \mu_1 < \mu_2 < \cdots < \mu_r < \pi.$$

The restriction of the values μ_j to the interval $(0, \pi)$ is unessential, inasmuch as the multiplication of the jth column of equation (1) by $\exp(-\lambda A_j)$ leaves the type of the equation unchanged and substitutes $(-A_j)$ in the rôle of A_j. We redesignate now by A_{hk}, $k=1, 2, \cdots$, those of the quantities A_j with argument μ_h, arranging them in the order

$$|A_{hk}| < |A_{h,k+1}|.$$

The rays

$$\arg \lambda = \frac{\pi}{2} - \mu_h \tag{4}$$

proceed with h in clockwise succession in the λ plane, and are followed in this succession by the rays

$$\arg \lambda = -\frac{\pi}{2} - \mu_h. \tag{4'}$$

We shall confine the attention to a sector S_c enclosing a single one of the rays (4), (4'), i.e.

$$\arg \lambda = \pm \frac{\pi}{2} - \mu_c, \tag{5}$$

and defined by the relations

$$S_c: \quad \pm \pi/2 - \mu_c + \eta_c \geqq \arg \lambda > \pm \pi/2 - \mu_{c+1} + \eta_{c+1},$$

where η_c, η_{c+1} are positive constants chosen sufficiently small but otherwise arbitrary. The character of equation (1) in sector S_c is typical of its character

* Cf. Schwengeler, loc. cit., p. 48, and the final theorem of this note.

in any analogous sector associated with another of the rays (4) or (4′). The deductions for sector S_c, therefore, reveal the facts for the entire λ plane.

If the determinant in equation (1) is expanded the equation may be written in the form

$$\sum_{l=0}^{L} e^{\lambda K_l} \sum_{p=1}^{P} \omega_{l_p}(\lambda) e^{\lambda K_p'} = 0, \tag{6}$$

in which the constants K_p' are linear combinations with coefficients 0 and 1 of the quantities A_{hk}, $h \neq c$, and the constants K_l are similar combinations of the quantities A_{ck}. We may choose the subscripts, moreover, so that

$$0 = K_0 < |K_1| < \cdots < |K_L| .$$

Inasmuch as the quantities $R(\lambda A_{hk})$,* with $h \neq c$, maintain their signs throughout the sector S_c, the order by numerical magnitude of the quantities $R(\lambda K_p')$ is fixed for λ in the sector. If we suppose then that the subscript 1 is assigned so that

$$R(\lambda K_1') \geqq R(\lambda K_p') \qquad (p = 2, \cdots, P)$$

and divide equation (6) by $\exp(\lambda K_1')$ we have

$$\sum_{l=0}^{L} e^{\lambda K_l} \left\{ \omega_l(\lambda) + \sum_{p=2}^{P} \omega_{l_p}(\lambda) e^{\lambda H_p} \right\} = 0, \tag{7}$$

where $\omega_l = \omega_{l_1}$, and $H_p = K_p' - K_1'$. In S_c we have then

$$R(\lambda H_p) < 0. \tag{8}$$

Let l' and l'' be defined now by the relations

$$\omega_l \equiv 0 \text{ for } l < l', \text{ and } l > l'',$$
$$\omega_{l'} \not\equiv 0, \quad \omega_{l''} \not\equiv 0,\dagger$$

and let the sector S_c be divided into sub-sectors S_{c1} and S_{c2} as follows:

$$S_{c1}: \pm \pi/2 - \mu_c + \eta_c \geqq \arg \lambda > \pm \pi/2 - \mu_c - \eta_c,$$
$$S_{c2}: \pm \pi/2 - \mu_c - \eta_c \geqq \arg \lambda > \pm \pi/2 - \mu_{c+1} + \eta_{c+1}.$$

We observe then readily that if η_c is chosen sufficiently small we have in S_{c1} because of (8)

$$\{\omega_{l_p} e^{\lambda H_p}\} \exp(\lambda K_l) = \omega_{l_1}[0] \exp(\lambda K_{l_1}) \quad (p = 2, \cdots, P), \tag{9}$$

---

* The symbol $R(x)$ designates "the real part of x."

† In the simplest case we should have $l' = 0$, $l'' = L$.

for any values of l and l_1. Equation (7) may be written, therefore, in the form

$$\sum_{l=l'}^{l''} e^{\lambda K_l}\omega_l(\lambda)[1] = 0, \text{ in sector } S_{c1}. \tag{10}$$

In the sector S_{c2} on the other hand relations (9) are valid only for $l \leqq l_1$. We have, however, further in this case

$$\omega_l \exp(\lambda K_l) = \omega_{l_1}[0] \exp(\lambda K_{l_1}), \text{ for } l < l_1,$$

and hence we may write the equation after division by $\exp(\lambda K_{l''})$ in the form

$$\omega_{l''}(\lambda)[1] + \sum_{l=l''}^{L} \sum_{p=2}^{P} \omega_{l_p}(\lambda)[1] \exp(\lambda\{K_l - K_{l''} + H_p\}) = 0. \tag{11}$$

If $l'' = L$ the left member of this consists of its first term alone, and because of (3) there will be no roots of the equation in S_{c2} for $|\lambda|$ sufficiently large. On the other hand if $l'' < L$ the equation (11) is of precisely the original type (6) and we may repeat the foregoing discussion. The sector S_c and the ray (5) must be replaced in their rôles by the sector S_{c2} and the ray

$$\arg \lambda = \pm \pi/2 - \mu_{c2}, \tag{12}$$

where μ_{c2} is the largest value of the set

$$\arg\{K_l - K_{l''} + H_p\}, \ l > l'',$$

and the quantities λK_l must be replaced by those of the quantities $\lambda\{K_l - K_{l''} + H_p\}$ whose real part vanishes on the ray (12). We find from a repetition of the argument then that in a sector enclosing the ray (12) the equation is again of the general form (10), and that the remaining part of S_{c2} is either devoid of roots for $|\lambda|$ sufficiently large, or else that a further repetition of the procedure leads to a third ray to be considered. In any event the number of rays thus found in S_c and hence in the entire plane is finite. Every such ray is determined by a relation

$$R(\lambda Q) = 0, \tag{13}$$

where Q is formed by linear combination with coefficients 0, 1, and -1, from the values A_j of equation (1), and in an arbitrarily small sector enclosing such a ray the equation is of type (10), i.e. in virtue of (3) of the form

$$\sum_{m=0}^{M} \lambda^{r_m}[\alpha_m] e^{\lambda Q_m} = 0. \tag{14}$$

Excepting in these small sectors the equation has no roots for $|\lambda|$ sufficiently large.

The location of the roots of the equation (14) remains to be determined, and to this end we quote the following theorems which deal with the equation in certain specialized forms.

THEOREM A.* *If B and K are real, $B>0$, then for $|\rho|$ sufficiently large the roots ρ_m of the equation*

$$[a_1] + \rho^K[a_2]e^{\rho B} = 0 \tag{15}$$

are asymptotically spaced along the curve

$$\left| \rho^K\left(\frac{a_2}{a_1}\right)e^{\rho B} \right| = 1, \tag{16}$$

and

$$|\rho_m| \sim B^{-1}\left\{ \pm\left(2m - \frac{K}{2}\right)\pi + \arg\left(-\frac{a_2}{a_1}\right)\right\}.$$

Moreover if ρ remains uniformly away from the roots of (15) *then the left hand member of the equation is uniformly bounded from zero.*

We note that for $|\rho|$ sufficiently large one branch of the curve (16) lies entirely within an arbitrarily small sector about the positive axis of imaginaries and approaches parallelism with this axis, while the other branch is similarly located with respect to the negative axis of imaginaries.

THEOREM B.† *If the constants B_j are real and*

$$0 = B_0 < B_1 < \cdots < B_J,$$

then for $|\rho|$ sufficiently large the roots of the equation

$$\sum_{j=0}^{J}[b_j]e^{\rho B_j} = 0, \quad b_0 \neq 0, \quad b_J \neq 0, \tag{17}$$

lie in the strip bounded by the lines

$$R(\rho) = \pm c, \tag{18}$$

where c is a suitably chosen real constant. The number N of roots‡ lying in any interval of this strip of length l satisfies the relation

$$B_J l/(2\pi) - (J+1) \leqq N \leqq B_J l/(2\pi) + (J+1).$$

Moreover if ρ remains uniformly away from the zeros of the equation (17) *the left hand member of the equation is uniformly bounded from zero.*

---

* Wilder, loc. cit., pp. 424, 425. Tamarkin, loc. cit. (1), pp. 284–289. This theorem and Theorem B are proved by the authors cited under the hypotheses that K is an integer and that $[a]=a+E(\rho)/\rho$, where $E(\rho)$ is bounded for $|\rho|$ sufficiently large. Their proofs, however, are still valid with the more general hypotheses of this paper.

† Wilder, loc. cit., p. 421–423. Tamarkin, loc. cit. (1), pp. 168–173, (2), pp. 27–29.

‡ Counted with their multiplicities.

We consider now the transformation of this theorem by means of the substitution

$$\rho = \mu + \sigma \log \mu, \tag{19}$$

in which σ is a real constant and the principal value of the logarithm is to be understood. The region into which (18) is transformed is bounded by the curves

$$| \mu^{\sigma} e^{\mu} | = e^{\pm c}. \tag{20}$$

It is readily shown to be a strip asymptotically of constant width $2c$, and, like the curve (16), approaching parallelism with the axes of imaginaries.

If we designate by ρ' and ρ'' the real and imaginary parts of ρ, with corresponding notation for μ, the relation (19) may be resolved into the equalities

$$\rho' = \mu' + \sigma \log | \mu |,$$

$$\rho'' = \mu'' + \sigma \arg \mu.$$

From the first of these we obtain, upon dividing it by $|\mu|$, restricting ρ to the region (18), and allowing $|\mu|$ to increase indefinitely, the result $\mu'/|\mu| \to 0$, i.e. $\arg \mu \to \pm \pi/2$. It follows readily that the portion $T_{R,l}$ of region (20) which lies between the arcs $|\mu| = R$, and $|\mu| = R+l$, is asymptotically congruent, as $R \to \infty$ with l fixed, to the corresponding portion of strip (18), and this in turn approaches conformity with a piece of the strip of length l. Moreover if μ_1 and μ_2 lie in $T_{R,l}$ and correspond to ρ_1 and ρ_2, then clearly

$$(\rho_2 - \rho_1) - (\mu_2 - \mu_1) \to 0, \quad \text{as} \quad R \to \infty.$$

Hence when R is sufficiently large any two points of $T_{R,l}$ at a distance greater than δ from each other correspond to points of strip (18) which are at a distance greater than $\frac{1}{2}\delta$ from each other. Inasmuch as a function $\epsilon(\rho)$ clearly becomes a function $\epsilon(\mu)$ we may state the following theorem in which we have again replaced the μ of the preceding discussion by ρ.

THEOREM C. *If $0 = B_0 < B_1 < \cdots < B_J$, and σ is real, then for $|\rho|$ sufficiently large the roots of the equation*

$$\sum_{j=0}^{J} [b_j] \{ \rho^{\sigma} e^{\rho} \}^{B_j} = 0, \quad b_0 \neq 0, \quad b_J \neq 0, \tag{21}$$

lie within the strip bounded by a pair of curves (20). *The number N of roots*

lying in a portion of this strip cut out by a pair of arcs $|\rho| = R$, *and* $|\rho| = R+l$, *where* l *is fixed, satisfies when* R *is sufficiently large the relation*

$$\frac{B_J l}{2\pi} - (J+1) \leqq N \leqq \frac{B_J l}{2\pi} + (J+1).$$

Moreover if ρ *remains uniformly away from the zeros of* (21) *then the left hand member of the equation is uniformly bounded from zero.*

Consider now the equation of type (14), i.e.

$$\sum_{m=0}^{M} \rho^{\nu_m} [a_m] e^{\rho C_m} = 0, \tag{22}$$

in which the constants C_m are real and $0 = C_0 < C_1 < \cdots < C_m$. Since multiplication of the equation by any power of ρ is permissible we may suppose $\nu_0 = 0$. Let σ_1 and m_1 be determined by the relations

$$\nu_m/C_m \begin{cases} \leqq \sigma_1, & \text{for } 1 \leqq m < m_1, \\ = \sigma_1, & \text{for } m = m_1, \\ < \sigma_1, & \text{for } m > m_1^*, \end{cases}$$

and choose η_1, so that for every $m \geqq 1$

either $$\nu_m/C_m = \sigma_1$$

or $$\nu_m/C_m < \sigma_1 - \eta_1.$$

Further let S_1 designate the region bounded on the left by any ray through the origin with argument between $\pi/2$ and π, and on the right by the curve

$$|\rho^{\sigma_1} e^{\rho}| = e^{\eta_1}. \tag{23}$$

In the region S_1 we have

$$\rho^{\nu_m} e^{\rho C_m} = \begin{cases} \{\rho^{\sigma_1} e^{\rho}\}^{C_m}, & \text{if } \sigma_1 = \nu_m/C_m, \\ [0], & \text{if } \sigma_1 > \nu_m/C_m. \end{cases}$$

Hence in S_1 the equation (22) is of the form (15) or (21) and its roots are located as described in Theorem A or Theorem C.

On the other hand in any region S_2 bounded on the left by the curve (23) we have

$$\rho^{\nu_m} e^{\rho C_m} = \{\rho^{\sigma_1} e^{\rho}\}^{C_m} [0], \quad \text{for } m < m_1.$$

Hence equation (22) is of the form

---

* In the simplest case we should have $m_1 = M$.

$$\sum_{m=m_1}^{M} \rho^{\nu_m'}[a_m]e^{\rho C_m'} = 0, \quad \text{in} \quad S_2,$$

where $\nu_m' = \nu_m - \nu_{m_1}$, and $C_m' = C_m - C_{m_1}$. If $m_1 = M$ this latter equation has no roots for $|\rho|$ large. On the other hand if $m_1 < M$ the equation is precisely of the type (22) and the discussion applied above to S_1 may be repeated for S_2. The roots of equation (22) are thus found to be confined to a finite number of strips of the type (20).

Lastly to apply to equation (14) the results obtained for equation (22) it is necessary only to set

$$\lambda = \rho e^{-i \arg Q_m}.$$

In conclusion we state, therefore, the following theorem.

THEOREM. *For $|\lambda|$ sufficiently large the roots of equation* (1) *lie within arbitrarily small sectors each containing a ray* (13). *Within each of these sectors the roots are further confined to a finite number of strips which are asymptotically of constant width and approach parallelism with the ray* (13) *enclosed in the sector. The number of roots N, counted with their multiplicities, which lie in any portion of such a strip bounded by a pair of arcs $|\lambda| = R$, and $|\lambda| = R+l$, satisfies, for R sufficiently large, a relation*

$$\alpha l - \beta \leqq N \leqq \alpha l + \beta,$$

in which α and β are suitably determined constants. Moreover if λ remains uniformly away from a root of equation (1) *the left hand member of the equation is uniformly bounded from zero.*

UNIVERSITY OF WISCONSIN,
MADISON, WIS.

NETS WITH EQUAL W INVARIANTS*

BY

V. G. GROVE

1. Introduction

Let S_y be a non-developable surface in a projective space of three dimensions. On S_y let there be given two one-parameter families of curves such that one curve of each family passes through each point y of S_y, the two tangents being distinct. Such a set of curves will be called a *net* of curves. According to this agreement a conjugate net is a net. If, at any point of the discussion, the given net is conjugate, we shall explicitly call the net a conjugate net. Moreover we shall assume that the given net is not the asymptotic net.

The object of this paper is to generalize the property of isothermal-conjugacy in the sense that a certain property is to be defined for any non-asymptotic net, which property is isothermal-conjugacy if the net is conjugate. Without loss of generality we may assume that the given net has been made parametric. Let $y^{(k)}=y^{(k)}(u, v)$, $k=1, 2, 3, 4$, be the parametric equations of S_y. Let $Z^{(k)}=c^{(k)}$, $k=1, 2, 3, 4$, be the coördinates of a fixed point not lying in the tangent plane to S_y at y. Under these conditions the four functions y and the four constants c satisfy a system of differential equations of the form

$$
\begin{aligned}
y_{uu} &= a^{(11)}y_u + b^{(11)}y_v + c^{(11)}y + d^{(11)}z,\\
y_{uv} &= a^{(12)}y_u + b^{(12)}y_v + c^{(12)}y + d^{(12)}z,\\
y_{vv} &= a^{(22)}y_u + b^{(22)}y_v + c^{(22)}y + d^{(22)}z,\\
z_u &= z_v = 0.
\end{aligned}
\tag{1}
$$

The coefficients of system (1) satisfy the following integrability conditions:†

$$
\begin{aligned}
&a_u^{(12)} - a_v^{(11)} + a^{(12)}b^{(12)} - b^{(11)}a^{(22)} + c^{(12)} = 0,\\
&b_u^{(12)} - b_v^{(11)} + a^{(12)}b^{(11)} + (b^{(12)} - a^{(11)})b^{(12)} - b^{(11)}b^{(22)} - c^{(11)} = 0,\\
&c_u^{(12)} - c_v^{(11)} + a^{(12)}c^{(11)} + (b^{(12)} - a^{(11)})c^{(12)} - b^{(11)}c^{(22)} = 0,\\
&d_u^{(12)} - d_v^{(11)} + a^{(12)}d^{(11)} + (b^{(12)} - a^{(11)})d^{(12)} - b^{(11)}d^{(22)} = 0\ ;
\end{aligned}
\tag{2a}
$$

---

* Presented to the Society, September 8, 1927; received by the editors September 12, 1928, and (revised) May 28, 1929.

† G. M. Green, *Memoir on the general theory of surfaces and rectilinear congruences*, these Transactions, vol. 20 (1919), p. 150.

$$
\begin{aligned}
&a_u^{(22)} - a_v^{(12)} + a^{(22)}a^{(11)} + (b^{(22)} - a^{(12)})a^{(12)} - b^{(12)}a^{(22)} + c^{(22)} = 0,\\
&b_u^{(22)} - b_v^{(12)} + a^{(22)}b^{(11)} - a^{(12)}b^{(12)} - c^{(12)} = 0,\\
&c_u^{(22)} - c_v^{(12)} + a^{(22)}c^{(11)} + (b^{(22)} - a^{(12)})c^{(12)} - b^{(12)}c^{(22)} = 0,\\
&d_u^{(22)} - d_v^{(12)} + a^{(22)}d^{(11)} + (b^{(22)} - a^{(12)})d^{(12)} - b^{(12)}d^{(22)} = 0.
\end{aligned} \tag{2b}
$$

For convenience we adopt the following notation:

$$
\begin{aligned}
&\alpha = d^{(12)}/d^{(11)}, \quad \beta = a^{(12)} - \alpha a^{(11)}, \quad \gamma = b^{(12)} - \alpha b^{(11)},\\
&\qquad\qquad \delta = c^{(12)} - \alpha c^{(11)},\\
&\alpha' = d^{(12)}/d^{(22)}, \quad \beta' = a^{(12)} - \alpha' a^{(22)}, \quad \gamma' = b^{(12)} - \alpha' b^{(22)},\\
&\qquad\qquad \delta' = c^{(12)} - \alpha' c^{(22)};
\end{aligned} \tag{3}
$$

$$
\begin{aligned}
&a = d^{(11)}/d^{(22)}, \quad b = a^{(11)} - aa^{(22)}, \quad c = b^{(12)} - ab^{(22)}, \quad d = c^{(11)} - ac^{(22)},\\
&\qquad\qquad b' = a^{(12)}, \qquad\qquad c' = b^{(12)}, \qquad\qquad d' = c^{(12)}.
\end{aligned} \tag{4}
$$

We assemble at this point certain formulas which will be useful in our discussions:

$$
\begin{aligned}
K &= \delta + \gamma(\alpha\gamma + \beta) + \alpha\gamma_u - \gamma_v,\\
H' &= \delta' + \beta'(\alpha'\beta' + \gamma') + \alpha'\beta'_v - \beta'_u;
\end{aligned} \tag{5}
$$

$$
H = K - \beta'_u + \gamma_v, \quad K' = H' + \beta'_u - \gamma_v; \tag{6}
$$

$$
\begin{aligned}
&\mathfrak{D} = d + a\beta'^2 - \gamma^2 + a\beta'_v - \gamma_u + \beta' c + b\gamma + \alpha'(\beta'_u - \gamma_v),\\
&\alpha\mathfrak{D} = H' - K + (1 + \alpha\alpha')(\beta'_u - \gamma_v) = K - K' + \alpha\alpha'(\beta'_u - \gamma_v);
\end{aligned} \tag{7}
$$

$$
W^{(u)} = 2\gamma_v + (c/a)_u, \quad W^{(v)} = 2\beta'_u - b_v. \tag{8}
$$

From system (1) we note that the parametric net is conjugate if $d^{(12)} = 0$. Hence for conjugate nets we have $\alpha = \alpha' = 0$.

The functions (3) to (8) have been chosen to satisfy the following conditions:

(a) **Parametric net non-conjugate:** The functions α, β, γ, δ, α', β', γ', δ' are identical with the like named functions of Green's paper.* The functions K and H' are the same, except for a factor, as the functions h and k' respectively.† The functions $W^{(u)}$ and $W^{(v)}$ are the negatives of like named functions of Green's paper just cited.‡

---

* G. M. Green, *Nets of space curves*, these Transactions, vol. 21 (1920), p. 121. Hereafter referred to as Green, *Nets*.

† Green, *Nets*, pp. 216–217.

‡ Ibid., p. 223.

(b) **Parametric nets conjugate:** The functions a, b, c, d, b', c', d' are the same as the like named coefficients of system (16) of Green's first memoir on conjugate nets.* The invariants $H=H'$, $K=K'$, $\mathfrak{D}$, $W^{(u)}$ and $W^{(v)}$ are the same as the functions of like notation in the theory of conjugate nets.

In the usual way, we may show that the developables of the ray congruence correspond to the net defined by equating the quadratic differential form

$$aHdu^2 - \mathfrak{D}dudv - K'dv^2 \tag{9}$$

to zero. The developables of the axis congruence correspond to the net defined by equating to zero the quadratic differential form

$$a(K + W^{(u)})du^2 - [\mathfrak{D} + \alpha'(W^{(v)} - W^{(u)} - 2\beta_u' + 2\gamma_v)]dudv - (H' + W^{(v)})dv^2. \tag{10}$$

The asymptotic net is defined by equating the form

$$adu^2 + 2\alpha'dudv + dv^2 \tag{11}$$

to zero.

We shall say that a net has *equal point invariants of the first kind if*

$$H - K = 0. \tag{12}$$

If we compute the harmonic invariant of the quadratic forms (9) and (11), we find that *the ray curves form a conjugate net if and only if the given net has equal point invariants of the first kind. Conjugate nets*, therefore, *with equal point invariants of the first kind have equal point invariants in the usual sense.*

The axis curves form a conjugate net if and only if the invariant

$$W^{(u)} - W^{(v)} + K - H \tag{13}$$

vanishes.

2. Nets with equal W invariants

We shall say that a net is *an I net* if

$$W^{(v)} - W^{(u)} = 0. \tag{14}$$

Evidently *if a net is a conjugate I net it is an isothermally conjugate net.* From (12) and (13) we observe that *if a net has two of the following properties it has the third also: its ray curves form a conjugate net; its axis curves form a*

* G. M. Green, *Projective differential geometry of one-parameter families of space curves and conjugate nets*, first memoir, American Journal of Mathematics, vol. 37 (1915), p. 221. [Hereafter referred to as Green, *Conjugate nets*, I.

conjugate net; it is an I net. If the net is conjugate, this theorem is the familiar theorem concerning isothermally conjugate nets.*

From (14) we obtain what might be called a generalized theorem of Demoulin and Tzitzéica: *If the tangents to the curves of a net form W congruences, the net is an I net.*

Let us call a net *a non-harmonic net of the first kind* if

$$\mathfrak{D} \neq 0, \quad \alpha(W^{(v)} - W^{(u)} - 2\beta_u' + 2\gamma_v) = 0. \tag{15}$$

It follows therefore that *a non-conjugate non-harmonic net of the first kind is an I net if and only if it has equal point invariants of the first kind.* A non-harmonic conjugate net is a non-harmonic net of the first kind.

If we compare the quadratic forms (9) and (10), we observe that *the only non-harmonic I nets of the first kind for which the axis curves coincide with the ray curves are those having equal point invariants of the first kind and the tangents to whose curves form W congruences.* If the given net is conjugate this theorem becomes the familiar theorem of Green† as corrected‡ by Wilczynski.

We shall call a net *a non-harmonic net of the second kind* if

$$\mathfrak{D} \neq 0, \quad \alpha'(W^{(v)} - W^{(u)} - 2\beta_u' + 2\gamma_v) \neq 0. \tag{16}$$

If we impose the condition that the corresponding coefficients of the quadratic forms (9) and (10) be proportional, we find that

$$\begin{aligned} H = H' &= \frac{(W^{(u)} + \beta_u' - \gamma_v)(\beta_u' - \gamma_v)}{W^{(v)} - W^{(u)} - 2\beta_u' + 2\gamma_v}, \\ K = K' &= \frac{(W^{(v)} - \beta_u' + \gamma_v)(\beta_u' - \gamma_v)}{W^{(v)} - W^{(u)} - 2\beta_u' + 2\gamma_v}, \\ \mathfrak{D} &= \alpha'(\beta_u' - \gamma_v). \end{aligned} \tag{17}$$

In this case the ray and axis curves coincide with the curves defined by the differential equation§

$$\begin{aligned} \alpha(W^{(u)} + \beta_u' - \gamma_v)du^2 - \alpha'(W^{(v)} - W^{(u)} \\ - 2\beta_u' + 2\gamma_v)dudv - (W^{(v)} - \beta_u' + \gamma_v)dv^2 = 0. \end{aligned} \tag{18}$$

---

* G. M. Green, *Projective differential geometry of one-parameter families of space curves, and conjugate nets on a curved surface*, second memoir, American Journal of Mathematics, vol. 38 (1916), p. 320. Hereafter referred to as Green, *Conjugate nets*, II.

† Green, *Conjugate nets*, II, p. 322.

‡ E. J. Wilczynski, *Geometrical significance of isothermal-conjugacy of a net of curves*, American Journal of Mathematics, vol. 42 (1920), pp. 211–221. Hereafter referred to as Wilczynski, *Nets*.

§ Green, *Nets*, p. 235. We call attention to the error in Green's theorem immediately above formulas (89). The words "and only if" should be deleted from the statement of the theorem.

A net shall be called *an harmonic* net if and only if the ray and axis tangents separate harmonically the tangents to the curves of the net. Hence the parametric net is harmonic if and only if

$$\mathfrak{D} = \alpha'(W^{(v)} - W^{(u)} - 2\beta_u' + 2\gamma_v) = 0. \tag{19}$$

An harmonic conjugate net is an harmonic net.*

The axis and ray curves of an harmonic net will coincide if and only if

$$H(H' + W^{(v)}) = K'(K + W^{(u)}). \tag{20}$$

It follows therefore that *if a net is an harmonic net with equal point invariants of the first kind, the axis and ray curves will coincide if and only if the given net is an I net.* Conversely *if a net is an harmonic net whose ray and axis curves coincide, then the net is an I net if and only if it has equal point invariants of the first kind.*

3. Equal point invariants of the second kind

A net shall be said to have *equal point invariants of the second kind if the invariant* $H-H'$ vanishes. Formulas (5) and (6) imply that if $H-H'$ vanishes then $K-K'$ vanishes. It follows also from (5) and (6) that *conjugate nets have equal point invariants of the second kind.* Equations (17) show that *non-harmonic nets of the second kind with coincident ray and axis curves have equal point invariants of the second kind. A non-conjugate I net with equal point invariants of the first and second kind is harmonic.* In the next section we shall characterize geometrically nets having equal point invariants of the second kind.

Let the parametric net be non-conjugate and have equal point invariants of the second kind. If use be made of (7), we find that the differential equations of the ray and axis curves may be written

$$\begin{aligned} &aHdu^2 - \alpha'(\beta_u' - \gamma_v)dudv - Kdv^2 = 0, \\ &a(K + W^{(u)})du^2 + \alpha'(W^{(u)} - W^{(v)} + \beta_u' - \gamma_v)dudv - (H + W^{(v)})dv^2 = 0, \end{aligned} \tag{21}$$

respectively. It follows therefore that *the tangents to the curves of the axis (ray) curves of a non-conjugate net with equal point invariants of the second kind will separate the tangents to the curves of the given net harmonically if and only if the axis (ray) curves form a conjugate system.* Conversely, we may show by means of (7), (9), (10), and (13) that *if the axis (ray) tangents of a non-conjugate net separate the tangents to the curves of the given net and asymptotic*

* Wilczynski, *Nets*, p. 215.

tangents harmonically, the given net has equal point invariants of the second kind. We may summarize our results in the following theorem:

If a non-conjugate net has any two of the following properties it has the third also: its ray (axis) curves form a conjugate net; its ray (axis) tangents separate the tangents to the curves of the net harmonically; it has equal point invariants of the second kind.

Suppose that the given net is a net of plane curves. It follows that

$$K + W^{(u)} = 0, \quad H' + W^{(v)} = 0.$$

Therefore *if a net of plane curves has equal point invariants of the first and second kind it is an I net.* In case the net is conjugate this theorem becomes the theorem of Green:* *a conjugate net consisting of plane curves is isothermally conjugate if and only if it has equal Laplace-Darboux invariants.* Moreover, if a net is a non-conjugate net of plane curves with equal point invariants of the first and second kinds it is harmonic.

4. The associate conjugate, anti-ray and anti-axis tangents

The associate conjugate net of a given net has been defined by Green† to be that net the tangents to whose curves are the double rays of the involution determined by the asymptotic tangents and the tangents to the curves of the given net. The associate conjugate net of the parametric net is defined by the differential equation

$$a du^2 - dv^2 = 0. \tag{22}$$

The associate conjugate net of the net (22) will be called the *second associate conjugate net* of the given parametric net. The differential equation of this net is

$$\alpha' du^2 + 2du dv + \alpha dv^2 = 0. \tag{23}$$

The second associate conjugate net of a given net coincides with the net if and only if the given net is conjugate. The harmonic invariant of the quadratic form (10) and the quadratic form appearing in the left member of (23) is

$$\alpha'(K - H') - 2\alpha'(K - H) + \mathfrak{D}. \tag{24}$$

In case the parametric net is non-conjugate, (24) may be written

$$(\alpha\alpha' - 1)(H - H').$$

---

* Green, *Conjugate nets*, II, p. 322.

† Green, *Nets*, p. 213, *Conjugate nets*, II, p. 313.

Hence *a necessary and sufficient condition that the axis tangents of a non-conjugate net separate the second associate conjugate tangents harmonically is that the given net have equal point invariants of the second kind.*

With Green* we may define the anti-ray net of any given net as that net, the tangents to whose curves are the harmonic conjugates of the ray tangents with respect to the tangents to the curves of the given net. The anti-ray net of the parametric net is defined by the equation

$$aH du^2 + \mathfrak{D} du dv - K' dv^2 = 0. \tag{25}$$

We may similarly define the anti-axis net of a given net. The differential equation of the anti-axis net of the parametric net is

$$a(K + W^{(u)})du^2 + [\mathfrak{D} + \alpha'(W^{(v)} - W^{(u)} - 2\beta_u' + 2\gamma_v)]dudv - (H' + W^{(v)})dv^2 = 0. \tag{26}$$

The double rays of the involution determined by the axis and associate conjugate tangents of the parametric net are determined by

$$a[\mathfrak{D} + \alpha'(W^{(v)} - W^{(u)} - 2\beta_u' + 2\gamma_v)]du^2 + 2a(H' - K + W^{(v)} - W^{(u)})dudv + [\mathfrak{D} + \alpha'(W^{(v)} - W^{(u)} - 2\beta_u' + 2\gamma_v)]dv^2 = 0. \tag{27}$$

The double rays of the involution determined by the anti-ray and associate conjugate tangents of the parametric net are determined by

$$a\mathfrak{D}du^2 + 2a(H - K')dudv + \mathfrak{D}dv^2 = 0. \tag{28}$$

Let the parametric net be a non-harmonic net of the first kind. In that case the jacobians (27) and (28) coincide if and only if

$$2(K' - K) + W^{(v)} - W^{(u)} = 0. \tag{29}$$

From (29) we have the theorem of Green and Wilczynski: *A non-harmonic conjugate net whose axis tangents, anti-ray tangents, and associate conjugate tangents form three pairs of an involution at every point is isothermally conjugate.* Moreover since a non-conjugate I net with equal point invariants of the first kind is harmonic, we may state the following theorem:

If a net is a non-harmonic net of the first kind with equal point invariants of the second kind then the pair of axis tangents, the pair of anti-ray tangents, and the pair of associate conjugate tangents form three pairs of the same involution if and only if the given net is isothermally conjugate.

Suppose that the parametric net is a non-harmonic net of the second

* Green, *Conjugate nets*, II, pp. 309, 310.

kind with equal point invariants of the second kind. Under these conditions the jacobians (27) and (28) may be written

$$\begin{aligned} \alpha' du^2 + 2du dv + \alpha dv^2 &= 0, \\ \alpha' du^2 - 2du dv + \alpha dv^2 &= 0, \end{aligned} \tag{30}$$

respectively. It follows therefore that *the axis tangents, the anti-ray tangents and the associate conjugate tangents of a non-harmonic net of the second kind with equal point invariants of the second kind cannot belong to the same involution.* The two nets (30) are the second associate conjugate net and the net obtained by harmonic reflection of the second associate tangents in the parametric tangents.

In the last two theorems, as is the case in the theory of conjugate nets, the axis tangents may be replaced by the anti-axis tangents, and the anti-ray tangents by the ray tangents.

MICHIGAN STATE COLLEGE,
EAST LANSING, MICH.

CONTRIBUTIONS TO THE GENERAL THEORY OF TRANSFORMATIONS OF NETS*

BY

V. G. GROVE

1. Introduction

Let there be given a surface S_y in projective space of three dimensions. Suppose that on S_y we have two one-parameter families of curves such that through each point of S_y there passes one curve of each family, the two tangents being distinct. Such a set of curves will be called a net N_y. Suppose that through each point y of S_y there passes a line g of a congruence G, such that the developables of the congruence G intersect S_y in the curves of the given net N_y. Let S_z be another surface in the same projective three-space, in one-to-one point correspondence to S_y, corresponding points lying on the lines g of G. The developables of G intersect S_z in a net N_z. If neither N_y nor N_z is a focal surface of G, the nets N_y and N_z have been called *nets in relation C* or *C transforms*.† If the nets N_y and N_z are conjugate nets in relation C, they are in the relation of a transformation F.

In case N_y and N_z are F transforms, the cross ratio C formed by the pair of corresponding points and the pair of focal points on the line joining them is a projective invariant.‡ In case N_y and N_z are C transforms, there exists, of course, a similar invariant,§ which we have denoted by R. It is the purpose of this paper to show that some of the theorems concerning the invariant C related to the transformation F are capable of generalization to the transformation C and the associated invariant R. We also present a generalization of the transformation Ω and its associated invariant H to our extended class of nets.

Without loss of generality we may assume that N_y and N_z are parametric. Let the parametric equations of S_y and S_z be

$$y^{(k)} = y^{(k)}(u,v), \quad z^{(k)} = z^{(k)}(u,v) \qquad (k = 1,2,3,4)$$

---

* Presented to the Society, September 6, 1928; received by the editors September 12, 1928, and (revised) May 28, 1929.

† V. G. Grove, *Transformations of nets*, these Transactions, vol. 30 (1928), p. 483. Hereafter referred to as Grove, *Transformations*.

‡ W. C. Graustein, *An invariant of a general transformation of surfaces*, Bulletin of the American Mathematical Society, vol. 30 (1928), pp. 122ff.

§ Grove, *Transformations*, p. 493.

respectively. The four pairs of functions (y, z) satisfy a system of differential equations of the form*

$$
\begin{aligned}
y_{uu} &= \alpha y_u + \beta y_v + py + Lz, \\
y_{uv} &= a y_u + b y_v + cy + Mz, \\
y_{vv} &= \gamma y_u + \delta y_v + qy + Nz, \\
z_u &= m y_u \qquad\quad + fy + Az, \\
z_v &= \qquad\quad n y_v + gy + Bz.
\end{aligned} \tag{1}
$$

We shall denote the coefficients of the differential equations which y and z satisfy, when their rôles in the formulation of system (1) are interchanged, by the corresponding primed letters. We shall refer the reader to the paper just cited for their values† in terms of the coefficients of system (1).

The coefficients α, β, γ, etc. satisfy certain integrability conditions of which we use the following:

$$
\begin{aligned}
m_v + a(m - n) - g &= Am, \\
n_u + b(n - m) - f &= Bn, \\
B_u - A_v + (n - m)M &= 0.
\end{aligned} \tag{2}
$$

From the integrability conditions, Green has shown that the following relation holds:

$$(a + \delta + B)_u = (b + \alpha + A)_v. \tag{3}$$

A similar formula exists for the primed letters arising when the rôles of y and z are interchanged. If we make the transformation

$$y = \lambda \bar{y}, \quad z = \mu \bar{z},$$

wherein

$$4\frac{\partial}{\partial u}\log\lambda = b + \alpha + A, \quad 4\frac{\partial}{\partial v}\log\lambda = a + \delta + B,$$

$$\mu = (mn)^{1/3},$$

---

* Grove, *Transformations*, p. 484.

† On account of the difficulty of printing the symbols used in the previous papers, we have changed our notation. In order to enable the reader to make the change readily, we give the following transformation scheme:

$$\begin{pmatrix} a^{(11)}, b^{(11)}, c^{(11)}, d^{(11)}; & a^{(12)}, b^{(12)}, c^{(12)}, d^{(12)}; & a^{(22)}, b^{(22)}, c^{(22)}, d^{(22)}; & m, p, q; & n', p', q' \\ \alpha, & \beta, & p, & L; & a, & b, & c, & M; & \gamma, & \delta, & q, & N; & m, f, A; & n, g, B \end{pmatrix}.$$

In case the letters of the first row are Greek, the corresponding letters below are primed.

the resulting system of differential equations satisfied by $\bar{y}$ and $\bar{z}$ will be such that

$$(4) \qquad a + \delta + B = \alpha + b + A = a' + \delta' + B' = \alpha' + b' + A' = 0.$$

We shall hereafter assume that the two systems are such that equations (4) are satisfied.

2. The geometric adjoined system of differential equations

The coördinates of the tangent planes to S_y and S_z at corresponding points y and z may be taken to be

$$Y = |\, y, y_u, y_v \,|\,, \quad Z = |\, z, z_u, z_v \,|$$

respectively. The system of differential equations of the form (1) satisfied by the four pairs of functions (Y, Z) has the coefficients $\alpha'', \beta'', \gamma''$, etc. defined by the formulas

$$\begin{aligned}
d\alpha'' &= B'LM + M(\alpha M - aL + M_u) - N(bL - \beta M + L_u) - A'LN, \\
d\beta'' &= A'LM + M(bL - \beta M + L_u) - L(\alpha M - aL + M_u) - B'L^2, \\
da'' &= B'M^2 + L(aN - \gamma M) - NL_v + M(bN - \delta M + M_v) - A'MN, \\
db'' &= A'M^2 + N(bL - \beta M) - LN_u + M(aL - \alpha M + M_u) - B'LM, \\
d\gamma'' &= B'MN + M(aN - \gamma M + N_v) - N(\delta M - bN + M_v) - A'N^2, \\
(5) \quad d\delta'' &= A'MN + M(\delta M - bN + M_v) - L(aN - \gamma M + N_v) - B'LN, \\
L'' &= -L/(mn), \quad M'' = -M/(mn), \quad N'' = -N/(mn), \\
dm'' &= nNL' - mMM', \qquad dt'' = nNM' - mMN', \\
ds'' &= mLN' - nML', \qquad dn'' = mLN' - nMM', \\
df'' &= -Am'' - Bs'', \qquad dg'' = -Bn'' - At'', \\
A'' &= -A', \qquad B'' = -B', \quad d = M^2 - LN.
\end{aligned}$$

The system with coefficients (5) will be called the *geometric adjoint* of system (1).

We have called nets N_y and N_z in relation L, or L transforms, if they are in relation C, and if the developables of the congruence of lines of intersection of corresponding tangent planes to the sustaining surfaces correspond to these nets. The conditions that N_y and N_z in relation C be L transforms are*

$$(6) \qquad mMN' - nM'N = nL'M - mLM' = 0.$$

Let us interpret Y and Z as points instead of planes and denote by N_Y

* Grove, *Transformations*, p. 487.

and N_Z the nets generated by them as y varies over S_y. We see from (5) that if *the nets N_y and N_z are in relation L, the nets N_Y and N_Z are in relation C.* Similarly we may verify that if N_Y and N_Z are in relation L, the nets N_y and N_z are in relation C. The transformations C and L present a complete duality among the elements involved.

Suppose the net N_Y has equal point invariants of the first kind.* We shall say in this case that N_y has *equal tangential invariants of the first kind.* The net N_y has equal tangential invariants of the first kind if and only if

$$(7) \qquad (a'' - \gamma''M''/N'')_u - (b'' - \beta''M''/L'')_v = 0.$$

If use be made of formulas (5),equation (7) may be written in the following form:

$$(8) \qquad (a - \gamma M/N)_u - (b - \beta M/L)_v - \frac{\partial^2}{\partial u \partial v} \log\left(\frac{L}{N}\right) = 0.$$

Hence *if N_y has equal point and equal tangential invariants of the first kind it follows that*

$$(9) \qquad \frac{\partial^2}{\partial u \partial v} \log\left(\frac{L}{N}\right) = 0.$$

Similarly *if N_z has equal point and equal tangential invariants of the first kind*

$$\frac{\partial^2}{\partial u \partial v} \log\left(\frac{L'}{N'}\right) = 0.$$

Suppose now that N_y and N_z are non-conjugate nets in relation L. It follows from (6) that

$$L'/N' = m^2L/(n^2N).$$

Hence *if one of the non-conjugate nets N_y and N_z in relation L has equal point and equal tangential invariants of the first kind, the other will have the same property only if L is the transformation† K_R.*

3. The transformation Ω_H. The invariant H

Suppose that N_y and N_z are in relation L. From (5) it follows that the focal planes of the line h of intersection of the tangent planes to S_y and S_z at y and z form with these tangent planes the cross ratio

$$(10) \qquad H = m''/n'' = nL'N/(mLN'),$$

* V. G. Grove, *Nets with equal W invariants*, in the present number of these Transactions, pp. 845–852.

† Grove, *Transformations*, p. 495.

whatever the character of the nets N_y and N_z, provided of course neither is asymptotic. We shall say that N_y and N_z are *in the relation of a transformation* Ω_H if they are in relation L and if

$$(\log H)_{uv} = 0. \tag{11}$$

*If N_y and N_z are conjugate nets in relation Ω_{-1} they are Ω transforms in the sense of Eisenhart.**

Consider now the product of the invariants† $R=m/n$ and H. This product will be unity if and only if

$$L/N = L'/N'. \tag{12}$$

We have therefore the theorems of Slotnick‡: *If two conjugate nets are in relation F, and the product of the invariants R and H is unity, the asymptotic nets on the sustaining surfaces correspond and conversely. And if one of two conjugate nets in relation F is isothermally conjugate, the other is also if the product of R and H is unity.*

Suppose N_y and N_z are non-conjugate and in relation L. From (6) we have

$$M'/N' = mM/(nN).$$

If $RH=1$, it follows from (12) that

$$M'/L' = mM/(nL) = nM/(mL).$$

Hence

$$m^2 - n^2 = 0.$$

Therefore *if the product of the invariants R and H of non-conjugate nets in relation L is unity, the nets are either radial transforms or K_{-1} and Ω_{-1} transforms.* In this case *the asymptotic net on $S_y(S_z)$ corresponds to the harmonic reflection of that net in the parametric tangents on $S_y(S_z)$.* Conversely *if the sustaining surfaces of non-conjugate nets in relation L are such that their asymptotics correspond to their harmonic reflections§ in the tangents to the curves of the given nets, the product of R and H is unity. The transformation L is therefore Ω_{-1}.*

---

* Eisenhart, *Transformations of Surfaces*, Princeton University Press, p. 134.

† Grove, *Transformations*, p. 493.

‡ M. M. Slotnick, *A contribution to the theory of fundamental transformations of surfaces*, these Transactions, vol. 30 (1928), p. 201. Hereafter referred to as Slotnick, *Transformations*.

§ If it is understood that the harmonic reflection of the asymptotic tangents in the tangents to the curves of a conjugate net are these asymptotic tangents, the word "non-conjugate" may be deleted from these statements.

4. Triads of nets

Let N_y and N_z be nets in relation C. Define the net N_η by the formula

$$\eta = \theta y_u + \phi y_v + z.$$

One may readily verify that N_η will be in relation C to N_y if and only if the functions ϕ and θ satisfy the following system of differential equations:

$$\theta\beta + \phi b + \phi_u - \phi(\theta L + \phi M + A) = 0,$$
$$\phi\gamma + \theta a + \theta_u - \theta(\phi N + \theta M + B) = 0.$$

Suppose that we have three nets N_y, N_z, N_η so related that N_y and N_z; N_z and N_η; N_η and N_y are C transforms in the pairs indicated, by means of different congruences G but the same congruences H. Such a triad of nets will be said to form *an H triad. If the nets of an H triad are conjugate, the H triad is an harmonic triad of conjugate nets in the sense of Slotnick.**

Let the transformations C transforming N_y into N_z; N_z into N_η; and N_η into N_y be denoted by C_1, C_3 and C_2 respectively. We shall speak of the transformation C_2 as the product of the transformations C_1 and C_3 and write

$$C_2 = C_1C_3.$$

The coördinates y, z, η of corresponding points of an H triad satisfy the equations of the form

$$\begin{aligned} z_u &= m_1y_u + f_1y + A_1z, & z_v &= n_1y_v + g_1y + B_1z\ ; \\ \eta_u &= m_3z_u + f_3z + A_3\eta, & \eta_v &= n_3z_v + g_3z + B_3\eta\ ; \\ \eta_u &= m_2y_u + f_2y + A_2\eta, & \eta_v &= n_2y_v + g_2z + B_2\eta. \end{aligned} \tag{13}$$

From equations (13) we may derive the following relations:

$$f_1/m_1 = f_2/m_2\ ;\quad g_1/n_1 = g_2/n_2\ ;\quad A_1 = -f_3/m_3;\quad B_1 = -g_3/n_3.$$

Hence *if neither of the transformations C giving rise to an H triad is radial, and if one of the nets of the H triad is conjugate, the other two nets of the triad have the same property.*

Let R_1, R_2, R_3 be the three invariants R of the corresponding transformations C_1, C_2, C_3 giving rise to an H triad. From the relations existing between the coefficients of equations (13), one may readily prove that $R_1R_3 = R_2$. If the transformations C_i are each also transformations L, we may prove that $H_1H_3 = H_2$, wherein H_1, H_2 and H_3 are the invariants H of the transformations C_1, C_2 and C_3 respectively.

We may define also a *C triad* as a set of three nets N_y, N_z and N_η such that

* Slotnick, *Transformations*, p. 195.

they are in relation C by means of the same congruence G but by different congruences H. We may prove that the invariants R_i and H_i of nets of a C triad satisfy the equation $R_1R_3=R_2$, and, if the transformations C_i are also transformations L, that $H_1H_3=H_2$. We may state our results as follows: *If the product of two transformations L is a transformation L the three nets in question form either a C or an H triad. In either case the invariants R and H of the product transformation are equal respectively to the products of the invariants R and the products of the invariants H of the two transformations.*

Other theorems similar in character to those arising in the theory of F transformations readily suggest themselves. We shall not enter into these details at present.

5. Ranges of nets in relation C

Consider a net N_z in relation C to the net N_y by means of a congruence G. Let N_ζ be any other net in relation C to N_y by means of the same congruence G. The net N_ζ is defined by the expression

$$\zeta = z - \lambda y, \tag{14}$$

where λ is a given arbitrary function of u and v. The one-parameter family of nets $N_{\zeta(h)}$ in relation C to N_y defined by

$$\zeta(h) = z - h\lambda y, \tag{15}$$

where h is a parameter independent of u and v, will be called *a range of nets* in relation C to N_y.

As may easily be verified, the line of intersection of the tangent planes to the surfaces S_y and S_ζ at corresponding points y and ζ intersect in the line joining the points

$$r' = r + \frac{\lambda P}{m-\lambda}y, \quad s' = s + \frac{\lambda P'}{n-\lambda}y, \tag{16}$$

where

$$r = y_u + fy/m, \qquad s = y_v + gy/n,$$
$$P = f/m + A - \lambda_u/\lambda, \quad P' = g/n + B - \lambda_v/\lambda.$$

Hence the tangent planes to S_y and $S_{\zeta(h)}$ intersect in the line joining

$$r_h' = r + \frac{h\lambda P}{m-h\lambda}y, \quad s_h' = s + \frac{h\lambda P'}{n-h\lambda}y. \tag{17}$$

Referred to the tetrahedron y, r, s, z the envelope of the lines joining the points (17) is the conic

$$[(m-n)x_1 - nPx_2 - mP'x_3]^2 + 4m(m-n)P'x_1x_3 = 0, \quad x_4 = 0. \tag{18}$$

This conic is tangent to the parametric tangents at the points

$$R = y_u + \left(\frac{f}{m} + \frac{nP}{m-n}\right)y, \quad S = y_v + \left(\frac{g}{n} + \frac{mP'}{n-m}\right)y. \tag{19}$$

Suppose that N_y is not conjugate. The points R and S will coincide with the points of intersection* of the line g' in Green's relation R to g if and only if

$$\begin{aligned} An + f + a(m-n) - n\lambda_u/\lambda &= 0, \\ Bm + g + b(n-m) - m\lambda_v/\lambda &= 0. \end{aligned} \tag{20}$$

By means of the integrability conditions (2), these equations reduce to

$$\frac{\partial}{\partial u}\log\frac{\lambda}{n} = \frac{\partial}{\partial v}\log\frac{\lambda}{m} = 0. \tag{21}$$

It follows therefore that

$$(\log R)_{uv} = 0,$$

and

$$\lambda = c'm = cn, \tag{22}$$

where c and c' are constants. Therefore R is a constant. The nets N_y and N_z are consequently in relation K_R where R is a constant. From (22) it follows that

$$\frac{m-\lambda}{n-\lambda} = \frac{m(1-c')}{n(1-c)}.$$

Hence *if the conic* (18) *is tangent to the parametric tangents at their intersection with the line* g' *in Green's relation* R *to* g, *the range of nets* $N_{\xi(\lambda)}$ *is a range of nets in relation* K_R *to* N_y *where* R *is a constant.*

In case N_y and $N_{\xi(\lambda)}$ are radial transforms, the lines of intersection of corresponding tangent planes are lines of a pencil with vertex

$$P'y_u - Py_v + (fP'/m - gP/n)y.$$

This point will be indeterminate if $P = P' = 0$. Under these conditions we have

$$\lambda_u/\lambda = f/m + A = n_u/n, \quad \lambda_v/\lambda = g/n + B = m_v/m. \tag{23}$$

Since the integrability condition of (23) is satisfied it follows that

$$\lambda = cm$$

where c is a constant.

* Grove, *Transformations*, p. 491.

MICHIGAN STATE COLLEGE,
EAST LANSING, MICH.

NOTE ON INTEGRO-q-DIFFERENCE EQUATIONS*

BY

C. RAYMOND ADAMS

Introduction. We concern ourselves here with a functional equation in which the unknown function $f(x)$ is subjected simultaneously to the q-difference operator and to an integral operator of Volterra type; it may be written

$$\text{(A)} \qquad \sum_{i=0}^{n} a_i(x) f(q^{n-i}x) = b(x) + \int_0^x K(x,\xi) f(\xi)\, d\xi \qquad (n \geqq 1),$$

and n will be spoken of as the *order* of the equation. Certain facts about the solutions of the equation of first order when $|q|$ is $\leqq 1$ are already known, thanks to the work of Picone, Popovici, Nalli, and Tamarkin.† It is not unnatural that these writers should have confined their studies to the real domain.

Our purpose in this paper is to investigate the existence of solutions both in the complex and real domains for the equation of order n, including the case of $|q|<1$ and the essentially different case of $|q|>1$. There being major differences in the situation when $b(x)$ is not identically zero and when it is, we reserve the symbol (A) for the case of $b(x) \not\equiv 0$ and employ the designation (B) for the equation when $b(x)$ is $\equiv 0$. The associated homogeneous q-difference equation, to which (B) reduces when $K(x, \xi)$ vanishes identically, we denote by (C).

COMPLEX DOMAIN

1. In this part of our work we assume that the functions $a_i(x)$ and $b(x)$ are analytic‡ at $x=0$ and that $K(x, \xi)$ is analytic at the place $(0, 0)$; thus we have

$$a_i(x) = a_{i0} + a_{i1}x + a_{i2}x^2 + \cdots \qquad (i = 0, 1, \cdots, n),$$

$$\text{(1)} \qquad b(x) = b_0 + b_1 x + b_2 x^2 + \cdots,$$

$$K(x,\xi) = k_{00} + k_{10}x + k_{01}\xi + \cdots$$

---

* Presented to the Society, March 29, 1929; received by the editor of the Bulletin in April, 1929, and transferred to the Transactions.

† Tamarkin, *On Volterra's integro-functional equation*, these Transactions, vol. 28 (1926), pp. 426–431; references to the other authors cited may be found here. This paper will be referred to hereafter as [T]. To Tamarkin's references may be added the recent note by Colombo, *Su una classe de equazioni integro-funzionali a limiti variabili*, Bollettino della Unione Matematica Italiana, vol. 7 (1920), pp. 80–85.

‡ Clearly it is no more general to allow the functions $a_i(x)$ and $b(x)$ to have poles at the origin.

for $|x|<R$, $|\xi|<R'$. The constant q is understood to be $\neq 0$ and of modulus $\neq 1$.

The characteristic equation* of (C) is then

$$\sum_{i=0}^{n} a_{i0}\rho^{n-i} = 0. \tag{2}$$

Corresponding to any root ρ_i which is finite, not zero, and not equal to a positive integral power of q times another root, there exists a solution of (C),

$$(x) = x^{r_i}g(x), \tag{3}$$

where

$$r_i = \frac{\log \rho_i}{\log q} \tag{4}$$

and $g(x)$ is analytic at the origin. It is natural therefore to seek solutions of (B) of the same nature.

Since r_i is not in general an integer, the function (3) is in general multiple-valued and so a single-valued function of position only on a Riemann surface with branch point at the origin and a suitable number, finite or infinite, of sheets. The integral of such a function from 0 to x, if r_i is not an integer, is thus independent of the path only if we specify the character of the path to the extent of giving the number of times it makes a circuit of the branch point. Such a specification having been made, *and we assume this to have been done*, the right-hand member of (B) is a single-valued function of position on this surface when $f(x)$ is the kind of function in question.

Under these conditions and if the real part of r is >-1, we readily find by substitution that to the root ρ_i of (2) there corresponds a formal solution of (B),

$$s(x) = x^{r_i}(s_0 + s_1x + s_2x^2 + \cdots). \tag{5}$$

Now r_i is uniquely defined only when determinations for arg ρ_i and arg q are chosen. For the equation (C) it is immaterial what determinations be selected; *for the equation* (B), *however, it is essential that determinations be selected which will make the real part of* $r>-1$. That this is always possible is easily verified; we shall assume such a choice to have been made.

If none of the numbers $q^i(i=0, 1, 2, \cdots)$ is a root of (2), the equation (A) is satisfied formally by a power series

$$s(x) = s_0 + s_1x + s_2x^2 + \cdots. \tag{6}$$

* Cf., e.g., Adams, *On the linear ordinary q-difference equation*, Annals of Mathematics, (2), vol. 30 (1929), pp. 195–205. This will be referred to as [A].

2. **The case of $|q|>1$.** When $|q|$ is >1 the convergence of the formal solution (5) can be established immediately by the methods employed by the author in [A].* Only slight and natural modification of the details of that work is necessary. In constructing the convergence proof it is desirable at first to restrict the radii of convergence of $a_i(x)(i=0, 1, \cdots, n)$ and the associated radii of convergence of $K(x, \xi)$ all to exceed 1; that this hypothesis is not essential may be shown as in [A]. We thus obtain the following

THEOREM 1. *When $|q|$ is >1 and the equation* (B) *is satisfied formally by a series of the type* (5), *that series converges in the neighborhood of $x=0$ and represents there an analytic solution of* (B). *The region of analyticity of the solution may be extended by repeated use of the equation* (B) *itself out to the nearest zero of $a_0(x)$ or nearest singularity of the functions $a_i(x)(i=0, 1, \cdots, n)$ and $K(x, \xi)$.*

As in [A] we may also establish

THEOREM 2. *When $|q|$ is >1 and the equation* (A) *is satisfied by a series of the type* (6), *that series converges at and in the neighborhood of $x=0$ and represents there an analytic solution of* (A). *The region of analyticity of the solution may be extended by repeated use of the equation* (A) *itself out to the nearest zero of $a_0(x)$ or nearest singularity of the functions $a_i(x)(i=0, 1, \cdots, n)$, $b(x)$, and $K(x, \xi)$.*

We shall not attempt to say what is the most general solution of either (B) or (A). It may be noted, however, (i) that *any linear combination, with constant coefficients, of particular solutions of* (B) *is a solution of* (B); and (ii) that *the sum of a particular solution of* (A) *and any such linear combination of solutions of* (B) *is a solution of* (A).

3. **The case of $|q|<1$.** Although Picone and Tamarkin confined themselves to the real domain, some of their results carry over immediately to the complex domain if the given functions in the equations studied are assumed analytic in the neighborhood of the origin. The following theorem may be inferred from the work of [T] and a simple generalization thereof;† it also follows in part from Picone's results.

---

* Cf. the third and seventh paragraphs of §2.

† [T] treats the equation

$$u(x) = f(x) + s(x)u[\theta_1(x)] + \int_{-a}^{x} K(x,\xi)u[\theta_2(\xi)]d\xi,$$

where $u(x)$ stands for the function to be determined. The results of §1 of [T] follow in the same manner if the lower limit of integration is 0. If we make this change and set $\theta_1(x)=qx$ and $\theta_2(x)=x$, the equation clearly reduces to an equation of first order of our type (A).

THEOREM 3. *If for* $|q|<1$ *the sum* $\sum_{i=0}^{n-1}|a_{i0}/a_{n0}|$ *is* <1, *the equation* (B) *has no solution bounded in the neighborhood of* $x=0$ *that does not vanish there identically and the equation* (A) *has a unique solution analytic in that neighborhood.*

Results for the equation (A) which are in certain respects more inclusive may be obtained by the method of successive approximations. We employ the following set of equations:

$$\sum_{i=0}^{n}a_i(x)f_1(q^{n-i}x) = b(x),$$

$$\sum_{i=0}^{n}a_i(x)f_m(q^{n-i}x) = b(x) + \int_0^x K(x,\xi)f_{m-1}(\xi)d\xi \quad (m = 2,3,\cdots). \tag{7}$$

It is now assumed that the equation (C) *has a complete set of solutions,*

$$F_i(x) = x^{r_i}(s_{i0} + s_{i1}x + s_{i2}x^2 + \cdots) \text{ for } |x| < R \quad (i = 1,2,\cdots,n) \tag{8}$$

with

$$D = \begin{vmatrix} F_1(qx) & F_2(qx) & \cdots & F_n(qx) \\ F_1(q^2x) & F_2(q^2x) & \cdots & F_n(q^2x) \\ \cdots & \cdots & \cdots & \cdots \\ F_1(q^nx) & F_2(q^nx) & \cdots & F_n(q^nx) \end{vmatrix}$$

$$= x^{r_1+r_2+\cdots+r_n}(d_0 + d_1x + d_2x^2 + \cdots) \qquad (d_0 \neq 0).$$

Such a set of solutions will exist if the roots of (2) are all finite and different from zero and if no one of them is equal to q^m times another, where m is a positive integer or zero; they may exist even if these conditions are not all fulfilled.* *We assume further that* $a_0(0)$ *is* $\neq 0$, *in which case there is no loss of generality in taking* $a_0(x)\equiv 1$.

The first equation of the set (7) will then be satisfied by

$$f_1(x) = \sum_{i=1}^{n}l_i(x)F_i(x)$$

The work of §1 of [T] may be extended at once, without modification of the method of proof, to cover the equation

$$u(x) = f(x) + \sum_{i=1}^{n}s_i(x)u[\theta_i(x)] + \int_{-x}^{x}K(x,\xi)u[\theta_{n+1}(\xi)]d\xi.$$

Tamarkin's σ is then replaced by the sum of the bounds of the absolute values of the functions $s_i(x)$ $(i=1, 2, \cdots, n)$ on the interval $-X\leqq x\leqq X$. From this generalization we may draw the conclusions indicated in Theorem 3 when the order of (A) is >1.

* Cf. [A].

if the functions $l_i(x)(i=1, 2, \cdots, n)$ are determined by the relations

$$l_i(x) = \Sigma\,(D_i/D),$$

where D_i represents the determinant obtained from D by replacing $F_i(q^j x)(j=1, 2, \cdots, n-1)$ by zero and $F_i(q^n x)$ by $b(x)$, and the operator Σ is defined by the equation*

$$\Sigma g(qx) - \Sigma g(x) = g(x).$$

But we have

$$(D_i/D) = x^{-r_i}(\theta_0 + \theta_1 x + \theta_2 x^2 + \cdots),$$

and *if* $|q^{r_i}|$ *is* >1, we may clearly employ for Σ the evaluation

$$\Sigma g(x) = -g(x) - g(qx) - g(q^2x) - \cdots$$

and obtain

$$\Sigma\,(D_i/D) = x^{-r_i}\phi_i(x),$$

where $\phi_i(x)$ is analytic for $|x|<R$. Thus *if each root of* (2) *is in absolute value* >1, we obtain a solution $f_1(x)$ which is analytic for $|x|<R$.

Defining $f_0(\xi)$ as zero, we have from (7)

$$\sum_{i=0}^{n} a_i(x)[f_m(q^{n-i}x) - f_{m-1}(q^{n-i}x)] = \int_0^x K(x,\xi)[f_{m-1}(\xi) - f_{m-2}(\xi)]d\xi \qquad (m = 2,3, \cdots).$$

For $m=2$ we have

$$f_{m-1}(\xi) - f_{m-2}(\xi) = f_1(\xi),$$

which is analytic for $|\xi|<R$. Now let R_1 denote any positive number less than the smaller of the two numbers R, R'; we then have

$$|f_1(\xi)| \leqq \sum_{i=1}^{n} |l_i(\xi)F_i(\xi)| < M \qquad \text{for } |\xi| \leqq R_1,$$

where M is a suitably chosen positive constant. Let us assume in general that $f_{m-1}(\xi) - f_{m-2}(\xi)$ is analytic and

$$|f_{m-1}(\xi) - f_{m-2}(\xi)| < M_1 \qquad \text{for } |\xi| \leqq R_1,$$

and let us examine $|f_m(x) - f_{m-1}(x)|$. As above we have

* Cf. Adams, *Note on the existence of analytic solutions of non-homogeneous linear q-difference equations, ordinary and partial*, Annals of Mathematics, (2), vol. 27 (1925), pp. 73–83.

$$f_m(x) - f_{m-1}(x) = \sum_{i=1}^{n} l_i'(x) F_i(x),$$

where

$$l_i'(x) = \Sigma (D_i'/D),$$

D_i' denoting the determinant obtained from D by replacing $F_i(q^j x)$ $(j=1, 2, \cdots, n-1)$ by zero and $F_i(q^n x)$ by $\int_0^x K(x, \xi)[f_{m-1}(\xi) - f_{m-2}(\xi)]d\xi$. For $|x| \leqq R_1$, $|\xi| \leqq R_1$, and K a suitably chosen positive constant, we have

$$|K(x,\xi)| < K,$$

and thus

$$|D_i'/D| < KM_1R_1 \,|\, x^{-r_i}(\theta_0' + \theta_1' x + \theta_2' x^2 + \cdots)\,|,$$

a bound in which only M_1 is dependent upon m. Employing for Σ the same evaluation as before, we clearly obtain for $l_i'(x)$ a function analytic for $0 \neq |x| \leqq R_1$ and satisfying the inequality

$$|l_i'(x)| < KM_1R_1 |\, x^{-r_i}\phi_i'(x)| \qquad \text{for } |x| \leqq R_1.$$

It follows that we have

$$|f_m(x) - f_{m-1}(x)| < KM_1R_1M' \qquad \text{for } |x| \leqq R_1,$$

where M' denotes a bound for $\sum_{i=1}^{n} |x^{-r_i}\phi_i'(x)F_i(x)|$.

Therefore the series

$$f_1(x) + [f_2(x) - f_1(x)] + [f_3(x) - f_2(x)] + \cdots \tag{9}$$

is dominated by the series

$$M + MKR_1M' + M(KR_1M')^2 + \cdots \qquad \text{for } |x| \leqq R_1.$$

Thus if R_1 be chosen so small that KR_1M' is <1, the series (9) converges uniformly for $|x| \leqq R_1$ and

$$\lim_{m\to\infty} f_m(x) \text{ exists uniformly for } |x| \leqq R_1.$$

Since $f_m(x)$ is analytic, the limit function, which we denote by $f(x)$, is also analytic for $|x| \leqq R_1$; that $f(x)$ satisfies the equation (A) is readily verified.

We have therefore proved

THEOREM 4. *If for* $|q|<1$ *and* $a_{00} \neq 0$ *the equation* (C) *has a full set of solutions* (8) *and each root of* (2) *is in absolute value* >1, *the equation* (A) *has a solution analytic at and in the neighborhood of* $x=0$.

It is readily shown that if the hypothesis of Theorem 3,

$$\sum_{i=1}^{n} \left| \frac{a_{i0}}{a_{n0}} \right| < 1,$$

be satisfied, each root of (2) is >1; obviously, however, the latter hypothesis may be fulfilled without the former.

Real domain

4. Naturally in the above discussion the independent variable x may be restricted to the real domain without essential change in the results. It should be pointed out, however, that under less stringent hypotheses for the known functions certain conclusions are immediate.

The case of $|q|<1$. Let the function $K(x, \xi)/a_n(x)$ be bounded in the region $0 \leqq x \leqq X$, $0 \leqq \xi \leqq X$ and if it has discontinuities let them be regularly distributed (cf. [T]); let the functions $b(x)/a_n(x)$ and $a_i(x)/a_n(x) (i=0, 1, \cdots, n-1)$ be continuous on the interval $0 \leqq x \leqq X$, with

$$\left| \frac{b(x)}{a_n(x)} \right| \leqq b \quad \text{and} \quad \sum_{i=0}^{n-1} \left| \frac{a_i(x)}{a_n(x)} \right| \leqq \sigma,$$

b and σ being positive constants. *Under these conditions* we have, by §1 of [T] and the generalization indicated above,

THEOREM 5. *If σ is <1 the equation* (B) *has no solution bounded on the interval $0 \leqq x \leqq X$ that does not vanish there identically, and the equation* (A) *has a unique solution continuous on that interval.*

Under the hypotheses (i) that $K(x, \xi)$ is continuous in the vicinity of $x=0$, $\xi=0$; (ii) that $b(x)$ is continuous in the vicinity of $x=0$ and, if discontinuous at the point, satisfies the condition

$$|b(x)| < B|x|^m,$$

B being a suitably chosen constant; and (iii) that the other known functions in the equation (A) are analytic at $x=0$, the method of §3 yields

THEOREM 6. *If for $a_{00} \neq 0$ the equation* (C) *has a full set of solutions* (8) *and each root of* (2) *is in absolute value $> |q|^m$, the equation* (A) *has a solution continuous in the neighborhood of $x=0$; if $b(x)$ is continuous at $x=0$, the solution is also.*

5. It may be remarked that all of the theorems of this paper are valid also for the equations obtained from (A) and (B) by replacing $f(\xi)$ by $f(q^j\xi)$, where j has any one of the values $1, 2, \cdots, n-1$.

BROWN UNIVERSITY,
PROVIDENCE, R. I.

THE EXPANSION PROBLEM IN THE THEORY OF ORDINARY LINEAR DIFFERENTIAL SYSTEMS OF THE SECOND ORDER*

BY
RUDOLPH E. LANGER

1. **Introduction.** The system composed of a linear differential equation of order n with boundary conditions applying at the end points of an interval is known to be expressible also as a system of n differential equations of the first order with boundary conditions of corresponding type. In particular, if the original system is of the form

$$u^{(n)}(x) + P_1(x,\rho)u^{(n-1)}(x) + \cdots + P_n(x,\rho)u(x) = 0,$$

$$\sum_{l=0}^{n-1}\{\alpha_{jl}u^{(l)}(a) + \beta_{jl}u^{(l)}(b)\} = 0 \qquad (j = 1,2,\cdots,n), \tag{1}$$

with the coefficient $P_i(x, \rho)$ a polynomial of degree i in ρ, then it may also be written in the specific form

$$u_j'(x) = \sum_{l=1}^{n}\{p_{jl}(x)\rho + q_{jl}(x)\}u_l(x),$$

$$\sum_{l=1}^{n}\{\sigma_{jl}u_l(a) + \tau_{jl}u_l(b)\} = 0 \qquad (j = 1,2,\cdots,n). \tag{2}$$†

Moreover, if in the original system (1) the coefficients of the boundary conditions are polynomials in the parameter, the same is also true when the system is written in the form (2).

The systems (1) and (2) are both adapted to a consideration of the associated boundary problems and both of these problems have been discussed under hypotheses of considerable generality.‡ An essential restriction, however, which has been imposed in all but a few investigations dealing with sys-

---

* Presented to the Society, September 7, 1928, and August 29, 1929; received by the editors in June, 1928, and May, 1929.

† Wilder, C. E., these Transactions, vol. 29 (1927), p. 497.

‡ Cf. in the case of system (1), Tamarkin 1. *Some General Problems of the Theory of Ordinary Linear Differential Equations*, etc., Petrograd, 1917 (in Russian), 2. Mathematische Zeitschrift, vol. 27 (1927), p. 1; and in the case of system (2), Birkhoff and Langer, Proceedings of the American Academy of Arts and Sciences, vol. 58 (1923), p. 51.

tems of specialized types, is that certain so-called conditions of "regularity" be fulfilled.*

The present paper is a discussion of the systems (1) and (2) of the second order, and a first purpose of the paper is the consideration of the boundary problems when certain of the major hypotheses of regularity are omitted. This weakening of the restrictions imposed in general involves a sacrifice of the convergence of the series expansions for arbitrary functions in terms of the solutions of the differential system. It is shown, however, that in a wide class of cases in which these expansions are non-convergent, they are nevertheless summable by suitable means.

Beyond this extension of the theories to irregular cases a second purpose of the paper is to equalize the extent of development of the theories of the two systems. While, on the one hand, the system (2) includes the system (1) and is somewhat more general, the investigations concerned with it have hitherto been confined to the case in which the coefficients of the boundary conditions are free from the parameter.† In the present discussion the coefficients are taken to be polynomials and in this respect the development of the theory of system (2) is brought to a par with that of the system (1).‡ In the matter of the expansibility of arbitrary functions in series of characteristic solutions, on the other hand, the theory of system (2) includes theorems asserting that n such functions may be simultaneously expanded with a single determination of coefficients, whereas in this respect the theory of system (1) has not hitherto been extended beyond the theorems on the expansibility of a single function. The simultaneous expansions associated with system (1) are included in the following considerations.

Representation of the differential system in matrix form is largely but not exclusively used as the most convenient scheme of notation. To obviate confusion between matrix and scalar quantities, however, the former are exclusively designated by script capitals. In the case of a matrix in which both columns are the same the designating letter is modified by a succeeding dot, while in the case of a matrix in which both rows are the same a preceding dot is used. Such matrices will be referred to as vectors. As a general rule

---

* The irregular system of type (1) and second order in which the differential equation is of a specialized form and the boundary conditions are free from the parameter has been considered by Stone, these Transactions, vol. 29 (1927), p. 23. Certain special irregular systems of the third order have been considered by Hopkins, these Transactions, vol. 20 (1919), p. 245, and by Ward, Annals of Mathematics, (2), vol. 26 (1924), p. 21, and these Transactions, vol. 29 (1927), p. 716.

† Cf. Birkhoff and Langer, loc. cit.

‡ Tamarkin 2, loc. cit.

the notation in so far as matrices are concerned is that used by Birkhoff and the author in the paper cited.*

2. **The differential system of type (1).** The explicit form of the system of type (1) to be considered is

$$(3)\qquad \begin{aligned} &u''(\xi) + \{p_{11}(\xi)\rho + p_{10}(\xi)\}u'(\xi) + \{p_{22}(\xi)\rho^2 + p_{21}(\xi)\rho + p_{20}(\xi)\}u(\xi) = 0,\\ &\alpha_{j1}(\rho)u'(a) + \alpha_{j0}(\rho)u(a) + \beta_{j1}(\rho)u'(b) + \beta_{j0}(\rho)u(b) = 0,\ j = 1,2, \end{aligned}$$

in which the α_{jl}, β_{jl} are any polynomials in ρ. The coefficients of the system may be real or complex. To obviate discussions not essential to the primary purpose of the paper, we shall suppose them to have derivatives of sufficiently high orders on the interval (a, b). As explicit hypotheses on the system (3), however, we shall assume as is usual† that on (a, b) the roots $\theta_1(\xi)$, $\theta_2(\xi)$, of the equation

$$\theta^2 + p_{11}(\xi)\theta + p_{22}(\xi) = 0,$$

are distinct, non-vanishing, and of constant argument, i.e.,

$$(4)\qquad \begin{aligned} &\theta_1(\xi) \not\equiv \theta_2(\xi),\\ &\theta_j(\xi) = |\,\theta_j(\xi)\,|\, e^{ic_j} \not\equiv 0,\ j = 1,2. \end{aligned}$$

If we suppose, as we may, that the arguments c_j are determined so that

$$|\,c_1 - c_2\,| \leqq \pi,$$

the change of variables and parameter

$$u = \psi(\xi)y, \quad x = (b - a)^{-1}(\xi - a), \quad \rho = \lambda e^{-i(c_1+c_2)/2},$$

with a suitable determination of the function $\psi(\xi)$, reduces the interval (a, b) to $(0, 1)$ and the system (3) to the form

$$(5)\qquad \begin{aligned} &\text{(a)}\ \ y''(x) - \{\lambda(r_1(x) + r_2(x)) + q_1(x)\}y'(x) + \{\lambda^2 r_1(x)r_2(x)\\ &\qquad\qquad + \lambda(r_1(x)q_1(x) - r_1'(x)) + q_2(x)\}y(x) = 0,\\ &\text{(b)}\ \ \omega_{j1}(\lambda)y'(0) + \omega_{j2}(\lambda)y(0) + \omega_{j3}(\lambda)y'(1) + \omega_{j4}(\lambda)y(1) = 0, \qquad j = 1,2. \end{aligned}$$

In this form, which we designate as the *normal form*, the coefficients $\omega_{jl}(\lambda)$ are again polynomials, and if we write

$$r_j(x) = \nu_j\,|\,r_j(x)\,|\,,$$

we have

* The reader is referred to that paper for a brief introductory exposition of the differential system in matrix form.

† Cf. Tamarkin 2, loc. cit., pp. 3 and 21; Birkhoff and Langer, loc. cit., pp. 72 and 109. Bliss, however, has considered the differential system with a set of conditions under which this hypothesis may not be fulfilled; these Transactions, vol. 28 (1926), p. 561.

$$(6)\quad \text{(a)}\quad |r_j(x)| > 0,\ j = 1,2;$$

$$\text{(b)}\quad \nu_1 = \nu_2^{-1} = e^{i\alpha}, \qquad 0 \leqq \alpha \leqq \pi/2 \quad (\alpha \text{ constant});$$

$$\text{(c)}\quad r_1(x) \not\equiv r_2(x).$$

The system (5) may now be expressed in the form of the differential system (2) by setting $y = y_1$, and writing

$$(7)\quad \begin{aligned} y_1' &= \lambda r_1(x) y_1 + e^{\int_0^x q_1 dx} y_2, \\ y_2' &= \lambda r_2(x) y_2 - q_2(x) e^{-\int_0^x q_1 dx} y_1. \end{aligned}$$

3. **The matrix form of system (2).** In the notation of matrices the system (2) takes the form

$$(8)\quad \begin{aligned} U'(\xi) &= \{\mathcal{P}(\xi)\rho + \mathcal{Q}(\xi)\}\, U(\xi), \\ \mathcal{W}_a\, U(a) + \mathcal{W}_b\, U(b) &= \mathcal{O}, \end{aligned}$$

the elements of any column of a matrix solution $U(\xi)$ constituting a solution of system (2). This system we shall consider for the case in which the elements of the matrices $\mathcal{W}_a$ and $\mathcal{W}_b$ are any polynomials in ρ, and we shall suppose as in §2 that the coefficients involved have derivatives of such orders as may be required. With the hypothesis, which we shall now make, that the roots $\theta_1(\xi)$, $\theta_2(\xi)$, of the determinant equation

$$\begin{vmatrix} p_{11}(\xi) - \theta & p_{12}(\xi) \\ p_{21}(\xi) & p_{22}(\xi) - \theta \end{vmatrix} = 0,$$

satisfy the conditions (4), we may without loss of generality suppose that in system (8), $p_{jl}(\xi) = \delta_{jl}\theta_j(\xi)$,* where $\delta_{jj} = 1$, $\delta_{jl} = 0$, $j \neq l$.

The functions $\psi_1(\xi)$, $\psi_2(\xi)$ may now be determined so that the change of variables and parameter

$$U(\xi) = (\delta_{ij}\psi_j(\xi))\Upsilon(\xi), \quad x = (b-a)^{-1}(\xi - a), \quad \rho = \lambda^{-i(c_1+c_2)/2},$$

reduces the system (8) to the normal form

$$(9)\quad \text{(a)}\quad \Upsilon'(x) - \{\mathcal{R}(x)\lambda + \mathcal{B}(x)\}\Upsilon(x) = \mathcal{O},$$

$$\text{(b)}\quad \mathcal{W}_{12}(\lambda)\Upsilon(0) + \mathcal{W}_{34}(\lambda)\Upsilon(1) = \mathcal{O}.$$

In this form the elements of the matrices $\mathcal{W}_{12}$ and $\mathcal{W}_{43}$ are again polynomials, while

---

* Cf. Birkhoff and Langer, loc. cit., p. 72. Throughout this paper the symbol δ_{jl} will be used exclusively to designate the "Kronecker delta" as in the text.

$$\mathcal{R}(x) = \begin{vmatrix} r_1(x) & 0 \\ 0 & r_2(x) \end{vmatrix}, \quad \mathcal{B}(x) = \begin{vmatrix} 0 & b_1(x) \\ b_2(x) & 0 \end{vmatrix},$$

and the functions $r_1(x)$, $r_2(x)$ satisfy the conditions (6).

We observe now that the pair of equations (7) are precisely in the form of the component equations of the vector system corresponding to (9), i.e., with

$$\text{(10)} \quad \begin{aligned} &\text{(a)} \qquad \Upsilon'(x)\cdot - \{\mathcal{R}(x)\lambda + \mathcal{B}(x)\}\Upsilon(x)\cdot = \mathcal{O}, \\ &\text{(b)} \qquad \mathcal{W}_{12}(\lambda)\Upsilon(0)\cdot + \mathcal{W}_{34}(\lambda)\Upsilon(1)\cdot = \mathcal{O}, \end{aligned}$$

with

$$\text{(11)} \qquad b_1(x) = e^{\int_0^x q_1(x)\,dx}, \quad b_2(x) = -q_2(x)e^{-\int_0^x q_1(x)\,dx}.$$

Hence the equation (5a) is equivalent to one of form (10a) in which the element $b_1(x)$ vanishes nowhere in the interval. Inasmuch as this property is invariant under a change of variables the greater generality of system (10) without restriction of $b_1(x)$ is evident. Any conclusions to be derived for the system (10) will be expressible in terms of the system (5) by means of the relations

$$\text{(12)} \quad \begin{aligned} &y_1(x) = y(x), && y(x) = y_1(x), \\ &y_2(x) = \{y'(x) - \lambda r_1(x)y(x)\}e^{-\int_0^x q_1\,dx}; && y'(x) = \lambda r_1(x)y_1(x) + e^{\int_0^x q_1\,dx}y_2(x). \end{aligned}$$

4. **The asymptotic solutions.** The asymptotic forms of certain solutions of the matrix equation (9a) and its adjoint equation

$$\text{(13)} \qquad \mathcal{Z}'(x) + \mathcal{Z}(x)\{\mathcal{R}(x)\lambda + \mathcal{B}(x)\} = \mathcal{O},$$

are known* when λ is confined to any region of the λ plane in which

$$\operatorname{Re}(\lambda r_1(x) - \lambda r_2(x))\dagger$$

either remains greater than or less than some constant for all x of the interval $(0, 1)$. To insure the availability of these asymptotic forms over the entire λ plane we make the usual hypothesis‡ that

$$\text{(14)} \qquad \arg\{r_1(x) - r_2(x)\} = A \qquad \text{(a constant)}.$$

This imposes no restriction on the functions $r_j(x)$ beyond those already involved in (6), when the argument α of (6b) is either 0 or $\pi/2$, for irrespective of the functions $r_j(x)$ the quantity (14) has in the former of these cases the

* Cf. Birkhoff and Langer, loc. cit., p. 87.

† The symbol Re(μ) will be used to designate "the real part of μ."

‡ Birkhoff and Langer, loc. cit., p. 83; Tamarkin 2, loc. cit., p. 21.

constant value 0 or π, as is evident from (6b) and (6c), while in the latter case it has the constant value $\pi/2$. When $0<\alpha<\pi/2$, however, the restriction imposed by the hypothesis (14) is clearly that the ratio $r_1(x)/r_2(x)$ be a constant.

We adopt now the familiar notation $[a(x)]$ to designate an expression of the form

$$[a(x)] = a(x) + \frac{a_1(x)}{\lambda} + \cdots + \frac{a_l(x)}{\lambda^l} + \frac{E(x,\lambda)}{\lambda^{l+1}}, \tag{15}$$

where l is any desired integer, and $E(x, \lambda)$ denotes a function which is analytic in λ and bounded for $|\lambda|$ sufficiently large. If we abbreviate then, by setting

$$R_j(x,t) = \int_t^x r_j(x)dx, \quad R_j(x) = R_j(x,0), \quad R_j = R_j(1,0),$$

the asymptotic solutions for the equations (9a) and (13), when λ remains in any half-plane bounded by a line parallel to the ray

$$\arg\lambda = \pi/2 - A,$$

are respectively as follows:

$$\mathcal{Y}(x) = \mathcal{P}(x,\lambda)\mathcal{E}(x,\lambda), \tag{16a}$$

where

$$e_{ij}(x,\lambda) = \delta_{ij}e^{\lambda R_j(x)}, \tag{16b}$$

$$p_{ij}(x,\lambda) = \lambda^{\delta_{ij}-1}[p_{ij}(x)], \quad p_{ii}(x) \equiv 1\ ; \tag{16c}$$

and

$$\mathcal{Z}(x) = \mathcal{E}^{-1}(x,\lambda)\mathcal{Q}(x,\lambda), \tag{17a}$$

$$q_{ij}(x,\lambda) = \lambda^{\delta_{ij}-1}[q_{ij}(x)], \quad q_{ii}(x) \equiv 1. \tag{17b}$$

Each column of the matrix (16a) is a solution of the vector equation (10a) and each row of the matrix (17a) is a solution of the vector equation

$$\cdot\mathcal{Z}'(x) + \cdot\mathcal{Z}(x)\{\mathcal{R}(x)\lambda + \mathcal{B}(x)\} = \mathcal{O}\ . \tag{18}$$

5. **The boundary conditions. The adjoint system.** The boundary conditions which together with the equation (18) yield the vector system adjoint to (11) are of the form

$$\cdot\mathcal{Z}(0)\,\mathcal{U} + \cdot\mathcal{Z}(1)\,\overline{\mathcal{U}} = \mathcal{O}\ , \tag{19}$$

the matrices $\mathcal{U}$ and $\overline{\mathcal{U}}$ being required to satisfy the conditions (i), that the rectangular array of their elements

$$\| v_{ij},\ \bar{v}_{ij} \| \text{ be of rank 2,} \tag{20}$$

and (ii), that

$$\mathcal{W}_{12}\mathcal{U} - \mathcal{W}_{34}\overline{\mathcal{U}} = \mathcal{O}.^* \tag{21}$$

To obtain these boundary conditions in terms of the elements of the matrices in (10b) we proceed as follows.

Let $\mathcal{W}_{jl}(\lambda)$ be defined as the matrix whose elements are those of the jth and lth columns of the array

$$\left\| \begin{matrix} w_{11}(\lambda) & w_{12}(\lambda) & w_{13}(\lambda) & w_{14}(\lambda) \\ w_{21}(\lambda) & w_{22}(\lambda) & w_{23}(\lambda) & w_{24}(\lambda) \end{matrix} \right\|. \tag{22}$$

From consistency of this definition with the notation of equation (10b), it follows that the elements of (22) are those of the matrices $\mathcal{W}_{12}$ and $\mathcal{W}_{34}$, and are, therefore, polynomials in λ. Moreover, our hypotheses must include the assumption that for all values of λ the array (22) is of rank two. We now designate by $\mathfrak{I}_{hl}$, the matrix whose elements are zero excepting that in the hth row and lth column which is unity, i.e.,

$$\mathfrak{I}_{hl} = (\delta_{ih}\delta_{lj}). \tag{23}$$

Then it is readily verified that we may write

$$\mathcal{W}_{12} = \mathcal{W}_{h,2+l}\mathfrak{I}_{1h} + \mathcal{W}_{m,2+n}\mathfrak{I}_{1m},$$
$$\mathcal{W}_{34} = \mathcal{W}_{h,2+l}\mathfrak{I}_{2l} + \mathcal{W}_{m,2+n}\mathfrak{I}_{2n}; \quad h,l = 1,2; \quad h+m = l+n = 3,$$

with any choice of the subscripts subject to the restrictions indicated. If $\Upsilon_{hl}\cdot$ and $\cdot\mathrm{Z}_{hl}$ are then defined by the relations

$$\begin{aligned} &\Upsilon_{12}\cdot = \Upsilon(0)\cdot, \quad \Upsilon_{34}\cdot = \Upsilon(1)\cdot, \\ &\Upsilon_{hl}\cdot = \mathfrak{I}_{1h}\Upsilon(0)\cdot + \mathfrak{I}_{2,l-2}\Upsilon(1)\cdot, \text{ for } h = 1,2;\ l = 3,4; \end{aligned} \tag{24}$$

and

$$\begin{aligned} &\cdot\mathrm{Z}_{12} = -\cdot\mathrm{Z}(0), \quad \cdot\mathrm{Z}_{34} = \cdot\mathrm{Z}(1), \\ &\cdot\mathrm{Z}_{hl} = -\cdot\mathrm{Z}(0)\mathfrak{I}_{h1} + \cdot\mathrm{Z}(1)\mathfrak{I}_{l-2,2}, \quad \text{for} \quad h = 1,2;\ l = 3,4; \end{aligned} \tag{25}$$

it follows that

$$\mathcal{W}_{12}\Upsilon(0)\cdot + \mathcal{W}_{34}\Upsilon(1)\cdot \equiv \mathcal{W}_{hl}\Upsilon_{hl}\cdot + \mathcal{W}_{mn}\Upsilon_{mn}\cdot, \tag{26}$$

and furthermore that

* Bliss, loc. cit., p. 562.

$$\cdot Z(x)Y(x)\cdot \Big|_0^1 = \cdot Z_{hl}Y_{hl}\cdot + \cdot Z_{mn}Y_{mn}\cdot . \tag{27}$$

In these relations the subscripts may be chosen as any permutation of the numbers 1, 2, 3, 4, subject to $h<l$ and $m<n$.

Now for any value of λ the subscripts h, l may be so chosen that $\mathcal{W}_{hl}^{-1}(\lambda)$ exists. Then condition (10b) is found with the use of (26) to reduce the matrix (27) to the zero matrix if the vector $\cdot Z(x)$ satisfies the condition

$$\cdot Z_{mn} - \cdot Z_{hl}\mathcal{W}_{hl}^{-1}\mathcal{W}_{mn} = \mathcal{O}.$$

This relation expressed in the form (19), as it may be by the use of (25), is easily shown to satisfy the conditions (20) and (21). Hence we may write the system (10) and its adjoint system in the corresponding forms

$$\begin{aligned} &\text{(a)} \quad Y'(x)\cdot - \{R(x)\lambda + B(x)\}Y(x)\cdot = \mathcal{O}, \\ &\text{(b)} \quad Y_{hl}\cdot + \mathcal{W}_{hl}^{-1}(\lambda)\mathcal{W}_{mn}(\lambda)Y_{mn}\cdot = \mathcal{O}, \quad h<l\,;\ m<n\,; \end{aligned} \tag{28}$$

and

$$\begin{aligned} &\text{(a)} \quad \cdot Z'(x) + \cdot Z(x)\{R(x)\lambda + B(x)\} = \mathcal{O}, \\ &\text{(b)} \quad \cdot Z_{mn} - \cdot Z_{hl}\mathcal{W}_{hl}^{-1}(\lambda)\mathcal{W}_{mn}(\lambda) = \mathcal{O}, \quad h<l\,;\ m<n. \end{aligned} \tag{29}$$

6. **The characteristic values.** The characteristic equation whose roots are the values of λ for which the system (10) admits of a solution other than the trivial solution $Y(x)\cdot \equiv \mathcal{O}$, is obtained by equating to zero the determinant of the matrix on the left of equation (9b).* With the form (16a) substituted for $Y(x)$, this equation may be written

$$D(\lambda) = 0, \tag{30}$$

where

$$D(\lambda) \equiv \left| \mathcal{W}_{12}(\lambda)P(0,\lambda) + \mathcal{W}_{34}(\lambda)P(1,\lambda)E(1,\lambda) \right| . \tag{31}$$

If the values given by (16b) and (16c) are substituted in this expression for $D(\lambda)$, the latter becomes explicitly

$$D(\lambda) \equiv \begin{vmatrix} d_{11}(\lambda) + d_{13}(\lambda)e^{\lambda R_1} & d_{12}(\lambda) + d_{14}(\lambda)e^{\lambda R_2} \\ d_{21}(\lambda) + d_{23}(\lambda)e^{\lambda R_1} & d_{22}(\lambda) + d_{24}(\lambda)e^{\lambda R_2} \end{vmatrix}, \tag{32}$$

where we have abbreviated by setting

$$\begin{aligned} d_{ij}(\lambda) &= w_{i1}(\lambda)p_{1j}(0,\lambda) + w_{i2}(\lambda)p_{2j}(0,\lambda), \\ d_{i,j+2}(\lambda) &= w_{i3}(\lambda)p_{1j}(1,\lambda) + w_{i4}(\lambda)p_{2j}(1,\lambda), \quad i,j = 1,2. \end{aligned} \tag{33}$$

* Birkhoff and Langer, loc. cit., p. 87.

For convenience in the discussion of equation (30), as well as for later reference, we define the determinants $D_r(\lambda)$ by the formulas

$$(34)\qquad \begin{aligned} D_0(\lambda) &= |d_{i1}, d_{i2}|\,, \quad D_1(\lambda) = |d_{i3}, d_{i2}|\,, \quad D_2(\lambda) = |d_{i1}, d_{i4}|\,, \\ D_3(\lambda) &= |d_{i3}, d_{i4}|\,, \quad D_4(\lambda) = |d_{i2}, d_{i4}|\,, \quad D_5(\lambda) = |d_{i3}, d_{i1}|\,. \end{aligned}$$

In expanded form the expression (32) is, then,

$$(35)\qquad D(\lambda) \equiv D_0(\lambda) + D_1(\lambda)e^{\lambda R_1} + D_2(\lambda)e^{\lambda R_2} + D_3(\lambda)e^{\lambda(R_1+R_2)},$$

and in this the determinant $D_r(\lambda)$, in virtue of (33) and (16c), either vanishes identically or else is of the form

$$D_r(\lambda) = \lambda^{\kappa_r}[\alpha_r],$$

with $\alpha_r \neq 0$ and κ_r an integer.

The conditions of "regularity" of the boundary problem to which reference has already been made, and which are usually imposed,* may be summarized here briefly as follows: If in (6b) $\alpha = 0$, then $D_0 D_3 \not\equiv 0$, and $\kappa_0 = \kappa_3 \geqq \kappa_1, \kappa_2, \kappa_4, \kappa_5$; if $\alpha = \pi/2$ then $D_1 D_2 \not\equiv 0$, and $\kappa_1 = \kappa_2 \geqq \kappa_0, \kappa_3, \kappa_4, \kappa_5$; and if $0 < \alpha < \pi/2$ then $D_r \not\equiv 0$, $r = 0, 1, 2, 3$, and $\kappa_0 = \kappa_1 = \kappa_2 = \kappa_3 \geqq \kappa_4, \kappa_5$. In the present discussion we shall omit entirely these restrictions on the values of the integers κ_r.

The transcendental equation $D(\lambda) = 0$, under variously restrictive hypotheses, has been discussed by Wilder, Tamarkin, Pólya and Schwengler. A discussion of the equation under hypotheses sufficiently general to include those of the present case has been given by the author,† and we shall therefore be content here merely to summarize the salient facts. If each coefficient $D_r(\lambda)$ in (35) is identically zero the characteristic equation (30) imposes no restriction, and the system (10) accordingly admits of a solution for every value of λ. On the other hand, if one and only one coefficient $D_r(\lambda)$ in (35) fails to vanish identically the characteristic equation can have no roots outside of a sufficiently large circle in the λ plane‡. In these cases there is no expansion problem. We shall not consider them further, but shall suppose in proceeding that at least two terms of $D(\lambda)$ actually occur.

* Tamarkin 2, loc. cit., pp. 23, 25; Birkhoff and Langer, loc. cit., p. 89.

† In the present issue of the Transactions, pp. 837–844.

‡ Examples of these cases are easily given. Thus, if in equation (9a) $r_1(x)$ and $r_2(x)$ are any distinct constants, and b_1 is a constant and $b_2 \equiv 0$, we have as an actual solution of the equation the matrix

$$Y(x) = \begin{pmatrix} e^{\lambda r_1 x} & \dfrac{\beta}{\lambda} e^{\lambda r_2 x} \\ 0 & e^{\lambda r_2 x} \end{pmatrix}, \quad \text{where} \quad \beta = b_1/(r_2 - r_1).$$

Then with the boundary condition

$$W_{12}(\lambda) = \begin{pmatrix} \lambda & -\beta \\ 0 & 0 \end{pmatrix}, \quad W_{34}(\lambda) = \begin{pmatrix} 0 & 0 \\ -\lambda & \beta \end{pmatrix},$$

In the case in which the constant α in (6b) is zero, the characteristic values lie in strips of asymptotically constant width which for $|\lambda|$ sufficiently large lie within arbitrarily small sectors including the positive and negative axes of imaginaries. The bounding curves of the strips, moreover, approach parallelism with these axes, and if l is any constant, then the number N_l of characteristic values lying between the circles $|\lambda|=\Lambda$, and $|\lambda|=\Lambda+l$, satisfies, when Λ is sufficiently large, a relation of the form

$$al - b \leqq N_l \leqq al + b, \tag{36}$$

a and b being specific constants. The case in which $\alpha=\pi/2$ is similar to that in which $\alpha=0$, with the exception that the strips containing the characteristic values then lie in small sectors about the axes of reals and approach parallelism with these axes.

Finally in the case $0<\alpha<\pi/2$ the characteristic values lie asymptotically at regular intervals along specific curves which approach parallelism with certain rays in the λ plane determined by the value α. If in the expression (35) no term vanishes identically, these rays are given by the formula

$$\arg\lambda = \pm\pi/2 \pm \alpha, \tag{37}$$

and there is just one curve marking the asymptotic location of characteristic values in any small sector enclosing one of these rays. If, on the other hand, the terms in (35) are not all different from zero, some of the rays (37) will have no corresponding curves of characteristic values and certain other rays in the plane may have that property. Inasmuch as we shall not consider the expansion problem in these latter cases, we shall go no further into details.

It is convenient for the considerations which follow to assign subscripts to the characteristic values in the order of their numerical magnitude, i.e., so that

$$|\lambda_s| \leqq |\lambda_{s+1}| .$$

As is customary, we shall designate the solutions of the systems (10) and (29) corresponding to the characteristic value λ_s, by $\Upsilon^s(x)\cdot$ and $\cdot Z^s(x)$ respectively.

---

system (10) is found to admit of the solution

$$\Upsilon(x)\cdot = \begin{pmatrix} \beta e^{\lambda r_2 x} \\ \lambda e^{\lambda r_2 x} \end{pmatrix}\cdot ,$$

for every value of λ. On the other hand with the value $\mathcal{W}_{12}(\lambda)$ above and

$$\mathcal{W}_{34}(\lambda) = \begin{pmatrix} 0 & 1 \\ -\lambda & \beta \end{pmatrix},$$

the characteristic equation is

$$\lambda e^{\lambda(r_1+r_2)} = 0,$$

and hence is satisfied only for $\lambda=0$.

7. **The contours C_n.** It is clear from the discussion of the preceding section that regardless of the value of α the number of characteristic values which lie in any given annular region with center at $\lambda=0$, width l, and inner radius of sufficient magnitude, is subject to a relation of the type (36). Hence it is seen directly that it is always possible to construct in the λ plane an infinite sequence of closed contours C_n, $n=n_1, n_2, \cdots$, having the following properties:

(i) Every point of any contour C_n lies at a distance greater than a definite positive constant δ from any characteristic value;

(ii) The contour C_n cuts the rays (37) and the axes of reals and imaginaries at the distance Λ_n from the origin, where $\Lambda_n\to\infty$ as $n\to\infty$ over the values $n_1, n_2, \cdots$;

(iii) The contour C_n contains in its interior just n characteristic values;

(iv) The number of characteristic values between two consecutive contours is bounded, i.e., $n_{r+1}-n_r\leqq N$, for all r.

These contours C_n are conveniently thought of as circles with centers at $\lambda=0$ and respective radii Λ_n, and to avoid unessential complications we shall in the subsequent discussions consider them in that way.

A further fact which we note here for future reference is that the quotient of $D(\lambda)$ by any one of its terms as given in (35) is bounded uniformly from zero provided, firstly, that λ remains uniformly away from the characteristic values and, secondly, that $|\lambda|$ is sufficiently large.* Since by property (i) above this condition is satisfied when λ is confined to the contours described, it follows that any quotient obtained by dividing a term of (35) by $D(\lambda)$ remains uniformly bounded on the contours C_n.

8. **The formal expansion of an arbitrary vector.** Let $\mathfrak{F}(x)\cdot$ be a vector chosen arbitrarily with integrable elements $f_1(x)$ and $f_2(x)$. We consider formally the possibility of developing such a vector in a series of characteristic vectors $\Upsilon^s(x)\cdot$, and seek to determine the coefficients c_s in the relation

$$\mathfrak{F}(x)\cdot \approx \sum_{s=1}^{\infty} c_s \Upsilon^s(x)\cdot. \tag{38}$$

The symbol $\approx$ is used here to emphasize the fact that the equality is implied in a formal sense only. The following deductions are to be purely heuristic and hence the assumptions to be made in the course of the reasoning are not to be considered as essential. In a later section the rigorous consideration of the results to be obtained will be made.

* Langer, loc. cit.

If we abbreviate by designating the left-hand members of equations (28a) and (29a) by $\mathcal{L}(Y, \lambda)\cdot$ and $\cdot\mathcal{M}(Z, \lambda)$, respectively, i.e.,

$$\mathcal{L}(Y,\lambda)\cdot \equiv Y'(x)\cdot - \{\mathcal{R}(x)\lambda + \mathcal{B}(x)\}Y(x)\cdot,$$
$$\cdot\mathcal{M}(Z,\lambda) \equiv \cdot Z'(x) + \cdot Z(x)\{\mathcal{R}(x)\lambda + \mathcal{B}(x)\},$$

the relation

$$\cdot Z^s(x)\mathcal{L}(Y^r,\lambda_r)\cdot + \cdot\mathcal{M}(Z^s,\lambda_s)Y^r(x)\cdot = \mathcal{O}$$

is an identity in the values r and s. Upon integration we obtain from it the equation

$$(\lambda_s - \lambda_r)\int_0^1 \cdot Z^s\mathcal{R}Y^r\cdot dx + \cdot Z^s(x)Y^r(x)\cdot \Big|_0^1 = \mathcal{O}. \tag{39}$$

If $W_{hl}(\lambda)$ designates, now, the determinant of the corresponding matrix $\mathcal{W}_{hl}(\lambda)$, this determinant is a polynomial in λ, and if there are characteristic values for which the determinant vanishes they can be only finite in number. With the possible exception of a limited number of values of r and s, then, the final matrix in the left-hand member of (39) can be written, in virtue of (27), (28b) and (29b), in the form

$$\cdot Z^sY^r\cdot \Big|_0^1 \equiv \cdot Z^s_{hl}\left\{\mathcal{W}^{-1}_{hl}(\lambda)\mathcal{W}_{mn}(\lambda)\right\}_{\lambda=\lambda_r}^{\lambda=\lambda_s} Y^r_{mn}\cdot.$$

Moreover, if the subscripts h, l are chosen so that the degree of the determinant $W_{hl}(\lambda)$ is at least as high as that of any other determinant formed from the columns of the array (22), we may write

$$\frac{1}{\lambda_s - \lambda_r}\left\{\mathcal{W}^{-1}_{hl}(\lambda)\mathcal{W}_{mn}(\lambda)\right\}_{\lambda=\lambda_r}^{\lambda=\lambda_s} = \sum_{k=0}^{K}\frac{\lambda_s^k}{W_{hl}(\lambda_s)}\mathcal{Q}^k(\lambda_r), \tag{40}$$

where the elements of the matrix $\mathcal{Q}^k(\lambda)$ are rational in λ, and $K+1$ is at most as great as the degree of $W_{hl}(\lambda)$. If $\lambda_r \neq \lambda_s$ it follows from this that relation (39) may be expressed in the form

$$\int_0^1 \cdot Z^s\mathcal{R}Y^r\cdot dx + \cdot Z^s_{hl}\sum_{k=0}^{K}\frac{\lambda_s{}^k}{W_{hl}(\lambda_s)}\mathcal{Q}^k(\lambda_r)Y^r_{mn}\cdot = \mathcal{O}. \tag{41}$$

If we suppose in proceeding that the characteristic values are all simple and distinct from the zeros of $W_{hl}(\lambda)$, the relation (41) is valid except possibly for $s=r$. Hence on multiplying the equation (41) by c_r, the general coefficient in (38), and summing with respect to r we obtain formally with the use of (38) the relation

$$\int_0^1 \cdot Z^s \mathcal{R} \mathfrak{F} \cdot dx + \cdot Z_{hl}^s K(\lambda_s) \cdot = c_s \left\{ \int_0^1 \cdot Z^s \mathcal{R} Y^s \cdot dx + \cdot Z_{hl}^s \sum_{k=0}^{K} \frac{\lambda_s^k}{W_{hl}(\lambda_s)} \mathcal{Q}^k(\lambda_s) Y_{mn}^s \cdot \right\}. \tag{42}$$

In this we have abbreviated by setting

$$\sum_{r=1}^{\infty} c_r \mathcal{Q}^k(\lambda_r) Y_{mn}^r \cdot = \mathfrak{C}^k \cdot \qquad (k = 0,1,2, \cdots, K), \tag{43}$$

and

$$\sum_{k=0}^{K} \frac{\lambda_s^k}{W_{hl}(\lambda_s)} \mathfrak{C}^k \cdot = K(\lambda_s) \cdot . \tag{44}$$

The relation (42) involves only matrices in which all elements are the same. Hence if we suppose the matrix on the right to be different from $\mathcal{O}$, the equality of the general element on the left with the general element on the right determines the constant c_s.

It is true that the matrix $K(\lambda_s) \cdot$ apparently involves the constants c_r through the relations (44) and (45), and that therefore the determination of c_s by (42) is elusive. We may formally obviate the difficulty in this point by supposing that the constants c_r can be chosen to yield not merely relation (38) but also and simultaneously the relations (43), in which the $\mathfrak{C}^k \cdot$ are constant vectors chosen at pleasure. We shall proceed from this standpoint.

It was observed in §3 that the conclusions drawn for the vector system (10) are expressible also for the system (5) by means of the relations (12). Hence if the vector relation (38) is resolved into its components, and $y_1^s(x)$ and $y_2^s(x)$ are expressed in terms of $y^s(x)$ and $y^{s\prime}(x)$, we obtain from it the simultaneous formal expansions

$$\begin{aligned} f_1(x) &\approx \sum_{s=1}^{\infty} c_s y^s(x), \\ f_2(x) &\approx \sum_{s=1}^{\infty} c_s \left\{ y^{s\prime}(x) - \lambda r_1(x) y^s(x) \right\} e^{-\int_0^x q_1(x)\,dx}, \end{aligned} \tag{45}$$

associated with the system (5). In these relations $f_1(x)$ and $f_2(x)$ are arbitrary, while the coefficients c_s are the same in both series and are determined by (42).

9. **The formal derivation of the contour integral.** The characteristic values of the adjoint systems (28) and (29) are the same, and if λ is not such a value there exists for the systems a Green's matrix $\mathcal{G}(x, t, \lambda)$ in terms of which the solution of the non-homogeneous system corresponding to (29), i.e.,

$$\begin{aligned} &\text{(a)} \quad \cdot \mathcal{M}(\mathcal{U}, \lambda) = \cdot \mathcal{A}(x), \\ &\text{(b)} \quad \cdot \mathcal{U}_{mn} - \cdot \mathcal{U}_{hl} \mathcal{W}_{hl}^{-1}(\lambda) \mathcal{W}_{mn}(\lambda) = \mathcal{O}, \end{aligned} \tag{46}$$

is given by the formula

$$\cdot \mathcal{U}(t) = -\int_0^1 \cdot \mathcal{A}(x)\, \mathcal{G}(x,t,\lambda)dx.^*$$

At the characteristic values the elements of $\mathcal{G}(x, t, \lambda)$ become infinite, and if, in particular, $\mathcal{G}(x,t,\lambda)$ has a simple pole at a given characteristic value the form of its residue there is known and is given by the formula

$$\mathcal{G}^s(x,t) = \lim_{\lambda\to\lambda_s} \mathcal{G}(x,t,\lambda)\{\lambda - \lambda_s\} = a_s \Upsilon^s(x) \cdot \cdot \mathcal{Z}^s(t), \tag{47}$$

where a_s is a specific constant.† With these facts at hand we shall seek formally to construct a function with residues which are the terms of the the series (38), and hence to express the sum of any number of terms of this series as a contour integral in the λ plane.

Let $\cdot \mathcal{C}(x, \lambda)$ be a vector to be subsequently determined, and let $\cdot \mathcal{U}(x, \lambda)$ be accordingly defined by the equation

$$\cdot \mathcal{U}(x,\lambda) = \cdot \mathcal{Z}^s(x) + \cdot \mathcal{C}(x,\lambda)\{\lambda - \lambda_s\}. \tag{48}$$

The substitution of $\cdot \mathcal{Z}^s(x)$ from this relation into the system (29) with $\lambda = \lambda_s$ yields for $\cdot \mathcal{U}(x, \lambda)$ the differential system

$$\begin{aligned} &\text{(a)} \quad \cdot \mathcal{M}(\mathcal{U},\lambda) = \{\lambda - \lambda_s\}\{\cdot \mathcal{U}(x,\lambda)\mathcal{R}(x) + \cdot \mathcal{M}(\cdot \mathcal{C},\lambda_s)\}, \\ &\text{(b)} \quad \cdot \mathcal{U}_{mn} - \cdot \mathcal{U}_{hl}\mathcal{W}_{hl}^{-1}(\lambda)\mathcal{W}_{mn}(\lambda) = \{\lambda - \lambda_s\} \cdot \mathcal{K}(\cdot \mathcal{C},\lambda_s,\lambda), \end{aligned} \tag{49}$$

where

$$\cdot \mathcal{K}(\cdot \mathcal{C},\lambda_s,\lambda) = \cdot \mathcal{C}_{mn}(\lambda) - \cdot \mathcal{C}_{hl}(\lambda)\mathcal{W}_{hl}^{-1}(\lambda)\mathcal{W}_{mn}(\lambda) - \cdot \mathcal{Z}_{hl}^s \sum_{k=0}^{K} \frac{\lambda_s{}^k}{W_{hl}(\lambda_s)} \mathcal{Q}^k(\lambda). \tag{50}$$

The relation (49b) is readily verified with the use of (40), and evidently reduces to the form (46b) if

$$\cdot \mathcal{K}(\cdot \mathcal{C},\lambda_s,\lambda) \equiv \mathcal{O}. \tag{51}$$

This identity (51), looked upon as a condition on the hitherto undetermined vector $\cdot \mathcal{C}(x, \lambda)$ contains but two component relations, and hence imposes but two conditions on the four elements of $\cdot \mathcal{C}(1, \lambda)$ and $\cdot \mathcal{C}(0, \lambda)$. It is clear, therefore, that $\cdot \mathcal{C}(x, \lambda)$ may always be chosen to satisfy (51). If we assume such a choice the system (49) may be formally considered as of the type (46), and hence we may write

$$\cdot \mathcal{U}(t,\lambda) = -\{\lambda - \lambda_s\}\int_0^1 \{\cdot \mathcal{U}(x,\lambda)\mathcal{R}(x) + \cdot \mathcal{M}(\cdot \mathcal{C},\lambda_s)\}\, \mathcal{G}(x,t,\lambda)dx.$$

* Birkhoff and Langer, loc. cit., pp. 67, 68.

† Birkhoff and Langer, loc. cit., p. 106.

As $\lambda \to \lambda_s$ this relation becomes in virtue of (47) and (48)

$$(52)\quad \cdot Z^s(t) = -a_s \int_0^1 \{ \cdot Z^s(x) R(x) Y^s(x) \cdot + \cdot M(\cdot C, \lambda_s) Y^s(x) \cdot \} dx \cdot Z^s(t).$$

The elimination of the vector $\cdot C$ from the equality (52) may now be made as follows. From equations (28) and (27) we find that

$$\int_0^1 \cdot M(\cdot C, \lambda_s) Y^s(x) \cdot dx = \cdot C(x, \lambda_s) Y^s(x) \cdot \Big|_0^1 = \cdot C_{hl}(\lambda_s) Y^s_{hl} \cdot + \cdot C_{mn}(\lambda_s) Y^s_{mn} \cdot$$
$$= \{ - \cdot C_{hl}(\lambda_s) W^{-1}_{hl}(\lambda_s) W_{mn}(\lambda_s) + \cdot C_{mn}(\lambda_s) \} Y^s_{mn} \cdot,$$

while in this final form the quantity within the brace is by (51) equal to the last term on the right of (50) with λ replaced by λ_s. Hence we may write (52) in the form

$$\cdot Z^s(t) = -a_s \Big\{ \int_0^1 \cdot Z^s R Y^s \cdot dx + \cdot Z^s_{hl} \sum_{k=0}^{K} \frac{{\lambda_s}^k}{W_{hl}(\lambda_s)} Q^k(\lambda_s) Y^s_{mn} \cdot \Big\} \cdot Z^s(t).$$

The coefficient of $\cdot Z^s(t)$ on the right of this relation is a matrix all of whose elements are the same, and because of this, as is easily verified, we may multiply the relation by $c_s Y^s(x) \cdot$ on the left and subsequently omit the the vector $\cdot Z^s(t)$ on the right. Then replacing the quantity within the brace by its equivalent as given by (42), we find that

$$(53)\quad c_s Y^s(x) \cdot = -a_s Y^s(x) \cdot \Big\{ \int_0^1 \cdot Z^s(t) R(t) \mathfrak{F}(t) \cdot dt + \cdot Z^s_{hl} K_{\cdot}(\lambda_s) \cdot \Big\}.$$

The vector $\cdot Z^s_{hl}$ by the definitions (25) is a linear combination of the vectors $\cdot Z^s(1)$ and $\cdot Z^s(0)$ with coefficients which are constant matrices. In virtue of formula (47), therefore, we may write (53) in the form

$$(54)\ c_s Y^s(x) \cdot = -\int_0^1 G^s(x,t) R(t) \mathfrak{F}(t) \cdot dt - \{ G^s(x,1) C_1 + G^s(x,0) C_0 \} K_{\cdot}(\lambda_s) \cdot,$$

where C_0 and C_1 are constant. Now the elements of $K_{\cdot}(\lambda) \cdot$ were observed in §8 to be rational functions of λ of degree at most (-1). We shall consider these rational functions expanded in powers of $1/\lambda$, and for $|\lambda|$ large we shall replace the elements of $K_{\cdot}(\lambda) \cdot$ by the leading terms of the resulting series. Then if C_n is a contour enclosing the first n characteristic values and no others we are led by (54) to the correspondence

$$(55)\quad \sum_{s=1}^{n} c_s Y^s(x) \cdot \sim \frac{-1}{2\pi i} \int_0^1 \int_{C_n} G(x,t,\lambda) R(t) \mathfrak{F}(t) \cdot d\lambda dt$$
$$- \frac{1}{2\pi i} \int_{C_n} \{ G(x,1,\lambda) \mathcal{A}_1 \cdot + G(x,0,\lambda) \mathcal{A}_0 \cdot \} \frac{d\lambda}{\lambda},$$

where the vectors $\mathcal{A}_1\cdot$ and $\mathcal{A}_0\cdot$ are constants, and where in using the symbol $\sim$ in place of the sign of equality we waive the questions of the terms in $K(\lambda)\cdot$ which have been dropped and also of the residue introduced by the integrand in this form at $\lambda=0$.

Let the vector $\mathcal{A}(x)\cdot$ be defined now by the formula

$$\mathcal{A}(x)\cdot = x\mathcal{A}_1\cdot + (x-1)\mathcal{A}_0\cdot .$$

Then recalling the discontinuity of the Green's matrix at the point $t=x$,* we find that

$$\begin{aligned} &G(x,1,\lambda)\mathcal{A}_1\cdot + G(x,0,\lambda)\mathcal{A}_0\cdot \\ &\quad = -\mathcal{A}(x)\cdot + \int_0^1 \{ G_t(x,t,\lambda)\mathcal{A}(t)\cdot + G(x,t,\lambda)\mathcal{A}'(t)\cdot \}\,dt \\ &\quad = -\mathcal{A}(x)\cdot - \int_0^1 G(x,t,\lambda)\{\lambda R(t)\mathcal{A}(t)\cdot + B(t)\mathcal{A}(t)\cdot - \mathcal{A}'(t)\cdot\}\,dt. \end{aligned}$$

It follows that the relation (55) may be written in the form

$$\begin{aligned} (56)\quad &\sum_{s=1}^{n} c_s Y^s(x)\cdot \sim \mathcal{A}(x)\cdot \\ &\quad -\frac{1}{2\pi i}\int_0^1 \int_{C_n} G(x,t,\lambda)\{R(t)\bar{\mathfrak{F}}(t)\cdot - \lambda^{-1}B(t)\mathcal{A}(t)\cdot + \lambda^{-1}\mathcal{A}'(t)\cdot\}\,d\lambda dt, \end{aligned}$$

where we have abbreviated by setting

$$(57)\qquad \mathfrak{F}(t)\cdot - \dot{\mathcal{A}}(t)\cdot = \bar{\mathfrak{F}}(t)\cdot .$$

The vector $K(\lambda)\cdot$ is given by (44) in terms of the vectors $\mathfrak{C}^k\cdot$ which we assumed in §8 could be chosen at pleasure. Since the coefficients of $\mathcal{A}(x)\cdot$ in turn depend upon $K\cdot$, it is clear that we may at this point consider these coefficients at our disposal. We shall choose them so that the vector (57) satisfies the conditions

$$(58)\qquad \bar{\mathfrak{F}}(0)\cdot = \mathcal{O}, \quad \bar{\mathfrak{F}}(1)\cdot = \mathcal{O},$$

namely,

$$(59)\qquad \mathcal{A}(x)\cdot = x\mathfrak{F}(1)\cdot + (1-x)\mathfrak{F}(0)\cdot .$$

Lastly we shall retain in the brace of the integrand in (56) only the term of highest degree in λ. This is consistent with our previous disposal of all but the term of highest degree in the vector $K\cdot$. The integral to which we are thus led is, then,

* Birkhoff and Langer, loc. cit., p. 70.

$$(60) \qquad \mathfrak{Z}_n(x,\mathfrak{F}\cdot) = \mathcal{A}(x)\cdot - \frac{1}{2\pi i}\int_0^1 \int_{C_n} \mathcal{G}(x,t,\lambda)\mathcal{R}(t)\bar{\mathfrak{F}}(t)\cdot d\lambda dt.$$

We shall consider in subsequent sections the convergence of this expression (60) as $n\to\infty$, the contours C_n being those described in §7. In that discussion we shall ignore the formal deductions above by which the expression $\mathfrak{Z}_n(x, \mathfrak{Z}\cdot)$ was deduced. This evidently means replacing the formal series (38) by the series of residues of the integrand in (60), and we shall henceforth refer to this series of residues as the expansion associated with the arbitrary vector $\mathfrak{F}(x)\cdot$. It will ultimately be shown that under suitable conditions this expansion is convergent.

10. **The contour integrals for the summability of the expansion.** The vector integral (60) may or may not converge as $n\to\infty$. Even in the case of non-convergence, however, its component integrals, or the corresponding expansions, may be summable by suitable means, and indeed we shall show later that such summability and its order depend directly upon the boundary conditions of the original differential system. The method of summation of series to which we shall resort is described as follows.

The components of the formal vector expansion (38) are given by the formulas

$$(61) \qquad f_i(x) \approx \sum_{r=1}^{\infty} c_r y_i^r(x) \qquad (i = 1,2).$$

In connection with them we shall consider the sequence of sums

$$(62) \qquad \sum_{r=1}^{n} c_r \left\{1 - \left(\frac{\lambda_r}{\Lambda_n}\right)^{\mu}\right\}^{\sigma_i} y_i^r(x) \qquad (i = 1,2\,;\, n = n_1, n_2, \cdots),$$

in which Λ_n is the radius of the contour C_n of §7 and μ is a positive constant whose precise specification we defer to a later section. The index σ_i is a constant, positive or zero, and for non-integral values of it we define the quantity with exponent σ_i by the formula

$$A^{\sigma_i} = e^{\sigma_i \log A},$$

where the principal value of the logarithm is to be understood.

The expressions (62) are a form of the Riesz typical means* for the series (61), and if for either value of i the sequence (62) converges the corresponding

* The application of the Riesz typical means in this manner to Birkhoff's and Fourier's series has been made by Stone, these Transactions, vol. 28 (1926), pp. 695–761; and vol. 29 (1927), pp. 25–53; also Bulletin of the American Mathematical Society, vol. 33 (1927), pp. 721–732. More detailed discussions and references are to be found in these papers.

expansion (61) is said to be summable of order σ_i. The extension of this definition to the vector form is obvious. Thus we shall define the matrix $\mathfrak{S}_n(\lambda)$ by the formulas

$$\mathfrak{S}_n(\lambda) = (\delta_{ij}s_{in}(\lambda)), \qquad s_{in}(\lambda) = \left\{1 - \left(\frac{\lambda}{\Lambda_n}\right)^{\mu}\right\}^{\sigma_i}, \tag{63}$$

and shall refer to the expansion (38) as summable of order (σ_1, σ_2), if the sequence of expressions

$$\sum_{r=1}^{n} c_r \mathfrak{S}_n(\lambda_r) Y^r(x) \cdot \tag{64}$$

converges as $n \to \infty$.

The deduction of a formulation for summability of the vector integral $\mathfrak{I}_n(x, \mathfrak{F}\cdot)$ of (60) is now easily made. Thus the general term of the expression (64) is obtainable by multiplying the relation (54) on the left by the matrix $\mathfrak{S}_n(\lambda_r)$. Then since $\mathfrak{S}_n(\lambda)$ is analytic over the entire plane the steps to the relations corresponding to (55) and (56) involve no difficulties, and yield the contour integral related to the sum (64). Precisely the considerations made at the close of §9 lead then to a study of the integral

$$\mathfrak{J}_n(x, \mathfrak{F}\cdot) = \mathcal{A}(x)\cdot - \frac{1}{2\pi i}\int_0^1 \int_{C_n} \mathfrak{S}_n(\lambda)\, \mathcal{G}(x,t,\lambda)\mathcal{R}(t)\bar{\mathfrak{F}}(t)\cdot d\lambda dt. \tag{65}$$

If this integral converges with particular values of the constants σ_i we shall say that the ith component of the integral (60) is summable of order σ_i, or, more briefly, that the expansion $\mathfrak{I}_n(x, \mathfrak{F}\cdot)$ is summable of order (σ_1, σ_2). Since $\mathfrak{S}_n(\lambda)$ reduces to the unit matrix for $\sigma_1 = \sigma_2 = 0$, the integral (60) is clearly included in the integral (65), and summability of order (0,0) signifies convergence.

11. **The simultaneous expansions associated with system (5).** The components of the formal vector expansion (38) were seen in §8 to yield the simultaneous formal expansions (45) associated with the differential system (5). The corresponding resolution of the vector relation (60) will in similar manner lead to a pair of contour integrals involving the arbitrary components $f_1(x), f_2(x)$, of the vector $\mathfrak{F}(x)\cdot$. The expression of these integrals in terms solely of quantities associated with the system (5) requires, to begin with, the determination of the relationship between the Green's function $G(x, t, \lambda)$ for system (5), and the elements of the Green's matrix $\mathcal{G}(x,t,\lambda)$ for the system (9).

If in system (7) the second equation is modified by the addition of a function $\phi(x)$ to its right-hand member, the resulting equations are the components of the vector equation

$$\Upsilon'(x)\cdot - \{\mathcal{R}(x)\lambda + \mathcal{B}(x)\}\Upsilon(x)\cdot = \begin{pmatrix} 0 \\ \phi(x) \end{pmatrix}\cdot, \tag{66}$$

with $b_i(x)$ and $q_i(x)$ related by formulas (11). On the other hand, if $y_2(x)$ is eliminated from this modified system (7), it is found that $y_1(x)$ satisfies the equation

$$l(y_1,\lambda) = \phi(x)b_1(x), \tag{67}$$

where $l(y, \lambda)$ is used to denote the left-hand member of equation (5a). Upon adjoining to equation (66) the boundary condition (10b), and to equation (67) the corresponding boundary conditions (5b), we obtain two non-homogeneous differential systems which are solved respectively by the formulas

$$\text{(a)} \qquad \Upsilon(x)\cdot = \int_0^1 \mathcal{G}(x,t,\lambda)\begin{pmatrix} 0 \\ \phi(t) \end{pmatrix}\cdot dt, \tag{68}$$

$$\text{(b)} \qquad y_1(x) = \int_0^1 G(x,t,\lambda)b_1(t)\phi(t)dt.$$

Inasmuch as the value of $y_1(x)$ given by the first component formula of (68a) must be identical with the value given by the formula (68b), regardless of the choice of the function $\phi(t)$, we conclude that

$$g_{12}(x,t,\lambda) = G(x,t,\lambda)b_1(t). \tag{69}$$

To obtain from this last equality the formulas for the remaining elements of the Green's matrix, we recall that in its points of continuity the matrix $\mathcal{G}(x, t, \lambda)$ satisfies the equation (9a) as a function of x, and satisfies the equation adjoint to (9a) as a function of t; i.e.,

$$\text{(a)} \qquad \mathcal{G}_x(x,t,\lambda) = \{\mathcal{R}(x)\lambda + \mathcal{B}(x)\}\,\mathcal{G}(x,t,\lambda), \tag{70}$$

$$\text{(b)} \qquad \mathcal{G}_t(x,t,\lambda) = -\,\mathcal{G}(x,t,\lambda)\{\mathcal{R}(t)\lambda + \mathcal{B}(t)\}.^*$$

From these equations we draw now respectively the component relations

$$\frac{\partial g_{12}}{\partial x} = \lambda r_1(x)g_{12} + b_1(x)g_{22},$$

$$\frac{\partial g_{12}}{\partial t} = -\lambda r_2(t)g_{12} - b_1(t)g_{11},$$

* Birkhoff and Langer, loc. cit., pp. 69, 70.

which yield the values of $g_{22}(x, t, \lambda)$ and $g_{11}(x, t, \lambda)$ in terms of the element (69). Lastly, a further component relation of (70a), i.e.,

$$\frac{\partial g_{11}}{\partial x} = \lambda r_1(x) g_{11} + b_1(x) g_{21},$$

yields the value of $g_{21}(x, t, \lambda)$ in terms of the value $g_{11}(x, t, \lambda)$ already at hand. For convenience in summarizing these results, we abbreviate by setting

$$\overline{G}(x,t,\lambda) = \{G_x(x,t,\lambda) - \lambda r_1(x) G(x,t,\lambda)\} e^{-\int_0^x q_1 dx}.$$

Then replacing the function $b_1(x)$ by its value as given in (11), we obtain the final formulas

$$\begin{aligned}
g_{11}(x,t,\lambda) &= -G_t(x,t,\lambda) - \{\lambda r_2(t) + q_1(t)\} G(x,t,\lambda), \\
g_{12}(x,t,\lambda) &= G(x,t,\lambda) e^{\int_0^t q_1 dt}, \\
g_{21}(x,t,\lambda) &= -\overline{G}_t(x,t,\lambda) - \{\lambda r_2(t) + q_1(t)\} \overline{G}(x,t,\lambda), \\
g_{22}(x,t,\lambda) &= \overline{G}(x,t,\lambda) e^{\int_0^t q_1 dt}.
\end{aligned} \tag{71}$$

With the relations thus deduced at hand the components of the integral (60) may now be expressed directly in terms of the differential system (5). Thus we find for them the formulas

$$\begin{aligned}
&\text{(a)} \quad I_{1n}(f_1,f_2) = a_1(x) + I_n(f_1,f_2,G), \\
&\text{(b)} \quad I_{2n}(f_1,f_2) = a_2(x) + I_n(f_1,f_2,\overline{G}),
\end{aligned} \tag{72}$$

where

$$I_n(f_1,f_2,G) = \frac{1}{2\pi i} \int_0^1 \int_{C_n} \Big\{ G_t(x,t,\lambda) r_1(t) \bar{f}_1(t) + G(x,t,\lambda) [\bar{f}_1(t) \{\lambda r_1(t) r_2(t) + r_1(t) q_1(t)\} - \bar{f}_2(t) r_2(t) e^{\int_0^t q_1 dt}] \Big\} d\lambda dt.$$

The convergence or summability of these expansions $I_{1n}(f_1, f_2)$ and $I_{2n}(f_1, f_2)$ to the respective values $f_1(x)$ and $f_2(x)$, will clearly follow from the argument which we shall develop in later sections to establish the convergence or summability of the vector integral (60) to the value $\mathfrak{F}(x)$.*

* The expansion theorem for a single arbitrary function in terms of the solutions of a system of type (5) is contained in the work of Tamarkin 2, loc. cit., §6, where the expansion formula is derived by methods distinct from those of the present paper. Despite a considerable difference in the final formulations, it is found that the expansion given by him for a function $f_1(x)$ is formally included in the first of the expansions (72) above, being deducible from the formula $I_{1n}(f_1, f_2)$ with the particular choice $f_2(x) = f_1'(x) \exp(-\int_0^x q_1 dx)$. The identification of the two results is most easily made by comparing the formulas of Tamarkin 2, pp. 35–36, with the first component relation of (54), the values (71) being substituted and suitable choice of the elements of the matrices $\mathcal{C}_0, \mathcal{C}_1$, and K being made. We note that the argument employed by Tamarkin can also be adapted for application to the differential system in matrix form. When applied in that manner it yields a derivation which may be considered alternative to that above for the expansion of an arbitrary vector.

12. **The Green's matrix.** The Green's matrix for the vector system (10) is given by the formula

$$\mathcal{G}(x,t,\lambda) = \tfrac{1}{2}\Upsilon(x)\{\pm \mathfrak{z} + \mathcal{D}^{-1}(\lambda)\{\mathcal{W}_{12}(\lambda)\Upsilon(0) - \mathcal{W}_{34}(\lambda)\Upsilon(1)\}\}\mathcal{Z}(t), \tag{73}$$

where the ambiguous sign is to be chosen + when $t<x$, and − when $t>x$.* In this formula the matrix $\mathfrak{z}$ is the unit matrix, $\Upsilon(x)$ and $\mathcal{Z}(x)$ are any pair of associated matrix solutions of the equation (9a) and its adjoint equation respectively, and $\mathcal{D}(\lambda)$ is the left-hand member of equation (9b). The substitution in formula (73) of the values of $\Upsilon(x)$ and $\mathcal{Z}(t)$ given by the relations (16) and (17) leads to the explicit evaluation of $\mathcal{G}(x,t,\lambda)$. We find it convenient to write the result of this computation in the form

$$\mathcal{G}(x,t,\lambda) = \sum_{l=0}^{3}\mathcal{G}^l(x,t,\lambda), \tag{74}$$

where, in terms of the matrices (23) and the determinants (31) and (34), we have the formulas

$$\text{(a)}\quad \mathcal{G}^0(x,t,\lambda) = \mathcal{P}(x,\lambda)\mathfrak{z}_{11}\mathcal{Q}(t,\lambda)D^{-1}(\lambda)\bar{D}(\lambda)e^{\lambda R_1(x,t)},$$

$$\text{where } \bar{D}(\lambda) = D_0(\lambda) + D_2(\lambda)e^{\lambda R_2} - \begin{cases}0 & \text{for } t \leqq x,\\ D(\lambda) & \text{for } t > x;\end{cases}$$

$$\text{(b)}\quad \mathcal{G}^1(x,t,\lambda) = \mathcal{P}(x,\lambda)\mathfrak{z}_{22}\mathcal{Q}(t,\lambda)D^{-1}(\lambda)\bar{\bar{D}}(\lambda)e^{\lambda R_3(x,t)},$$

$$\text{where } \bar{\bar{D}}(\lambda) = D_0(\lambda) + D_1(\lambda)e^{\lambda R_1} - \begin{cases}0 & \text{for } t \leqq x\\ D(\lambda) & \text{for } t > x;\end{cases} \tag{75}$$

$$\text{(c)}\quad \mathcal{G}^2(x,t,\lambda) = \mathcal{P}(x,\lambda)\mathfrak{z}_{12}\mathcal{Q}(t,\lambda)D^{-1}(\lambda)D_4(\lambda)e^{\lambda R_1(x)+\lambda R_2(1,t)};$$

$$\text{(d)}\quad \mathcal{G}^3(x,t,\lambda) = \mathcal{P}(x,\lambda)\mathfrak{z}_{21}\mathcal{Q}(t,\lambda)D^{-1}(\lambda)D_5(\lambda)e^{\lambda R_2(x)+\lambda R_1(1,t)}.$$

For convenience in substituting these terms in the integral (65), we note further that

$$\mathcal{S}_n(\lambda)\mathcal{P}(x,\lambda)\mathfrak{z}_{hl}\mathcal{Q}(t,\lambda)\mathcal{R}(t)\bar{\mathcal{F}}(t)\cdot = (s_{in}(\lambda)\lambda^{\delta_{ik}-1}\omega_{i,kl}(x,t,\lambda))\cdot, \tag{76}$$

where it is found from the formulas (16c) and (17b) that

$$\omega_{i,kl}(x,t,\lambda) = p_{ik}(x)r_l(t)\bar{f}_l(t) + \lambda^{-1}\{[\alpha^{(1)}_{i,kl}(x,t)]\bar{f}_1(t) + [\alpha^{(2)}_{i,kl}(x,t)]\bar{f}_2(t)\}, \tag{77}$$

the functions $\alpha^{(j)}_{i,kl}(x,t)$ being expressible in terms of the quantitites $[p_{ij}(x)]$ and $[q_{ij}(t)]$.

13. **The specification of the value μ.** The introduction of the factors $\{1-\alpha_n/\omega\}^\sigma$ into the partial sums

* Birkhoff and Langer, loc. cit., p. 110.

$$S_N(x) = \sum_{n=1}^{N} c_n y_n(x)$$

of an infinite series, to form the Riesz typical means

$$\sum_{n=1}^{N} \left\{1 - \frac{\alpha_n}{\omega}\right\}^{\sigma} c_n y_n(x), \qquad \alpha_N < \omega \leqq \alpha_{N+1},$$

is motivated by the purpose, usual in schemes of summation, of assigning varying weights to the sums S_N.* From this point of view the quantities α_n/ω are naturally specified by the rôle assigned them to be real and positive. Without adhering strictly to this viewpoint we shall nevertheless utilize it as a guide in the choice of the hitherto unspecified constant μ of formulas (62) and (63).

Consider first the case in which the constant α of (6b) is $\pi/2$. By §6 the characteristic values are confined then to strips of asymptotically constant width, bounded by curves which approach parallelism with the axes of reals. Hence for any sufficiently large value of j the quantity $\arg \lambda_j$ lies arbitrarily near one of the values 0 and π, and since Λ_n is real, it follows that as n and N become infinite, $n \leqq N$, the quantity $\{1-(\lambda_n/\Lambda_N)^{\mu}\}$ is asymptotically real and positive if μ is a multiple of 2. We shall, therefore, choose $\mu=2$ when $\alpha=\pi/2$.

If $\alpha=0$ the characteristic values, by §6, lie in strips which approach parallelism with the axes of imaginaries. Hence in this case the value $\arg \lambda_j$ for any j sufficiently large lies near one of the values $\pi/2$ and $3\pi/2$, and it follows that $\{1-(\lambda_n/\Lambda_N)^{\mu}\}$ is now asymptotically real and positive if μ is a multiple of 4. We are led, therefore, to the choice $\mu=4$ when $\alpha=0$.

In the case $0<\alpha<\pi/2$ the strips containing the characteristic values approach parallelism with certain other rays in the λ plane, these rays being four in number, and being given by the formulas (37) if no term of the characteristic equation (35) vanishes identically. We shall confine our considerations, when $0<\alpha<\pi/2$, to the case in which this hypothesis on the equation (35) is fulfilled.

If, in formula (63), $\sigma_i=0$ for either value of i, the element $s_{in}(\lambda)$ reduces to unity and the question of the value of μ is obviated so far as the corresponding component of (65) is concerned. This corresponds, of course, to the convergence of the respective component of the expansion (60). If $\sigma_i>0$, on

* Cf. Hardy and Riesz, *The General Theory of Dirichlet's Series*, Cambridge University Press, 1915, p. 22.

the other hand, μ must be chosen, and the principle followed in the previous cases clearly demands now a choice such that

$$\mu(\pi/2 \pm \alpha) \equiv 0 \pmod{2\pi}, \tag{78}$$

for both determinations of the ambiguous sign involved. It is found directly that the relations (78) are possible only if α is commensurable with π, and hence we shall restrict our further considerations when $\sigma_i > 0$ to this case.* The values of μ which satisfy (78) are then readily determined to be even integers, and we shall choose μ as the smallest positive value for which the relation (78) holds.

14. **Auxiliary theorems.** The proof of the convergence of the integral (65) as $n \to \infty$ is to be based upon certain classical theorems of Lebesgue and upon certain partial results which we shall summarize in the form of lemmas. For convenience in reference we group these theorems and lemmas together at this point.

Consider the integral

$$I(\psi, n) = \int_a^b \psi(z)\phi(z, n)dz,$$

where $\phi(z, n)$ is integrable with respect to z for the values n of a sequence $n_1, n_2, \cdots$; where $n_s \to \infty$, as $s \to \infty$. We quote the following two theorems.†

LEBESGUE THEOREM A. *A necessary and sufficient condition that*

$$\lim_{n\to\infty} I(\psi, n) = 0,$$

for every integrable function $\psi(z)$, is that

(i) $|\phi(z, n)| < M$, *a constant independent of z and n;*

(ii) $\lim_{n\to\infty} \int_{z_1}^{z_2} \phi(z, n)dz = 0$, *for* $a \leqq z_1 \leqq z_2 \leqq b$.

LEBESGUE THEOREM B. *A necessary and sufficient condition that*

$$\lim_{n\to\infty} I(\psi, n) = c\psi(a+),$$

for every function $\psi(z)$ which is of bounded variation on (a, b), is that

(i) $$\left| \int_{z_1}^{z_2} \phi(z, n)dz \right| < M, \quad \text{for} \quad a \leqq z_1 \leqq z_2 \leqq b,$$

* This hypothesis is fulfilled in the cases to which the "means" have been applied by Stone, loc. cit. It is, of course, also satisfied when $\alpha = 0$ or $\alpha = \pi/2$.

† Lebesgue, Annales de Toulouse, (3), vol. 1 (1909), p. 52 and p. 70.

(ii) $$\lim_{n\to\infty}\int_{z_1}^{z_2}\phi(z,n)dz = 0, \qquad for \quad a < z_1 \leqq z_2 \leqq b,$$

(iii) $$\lim_{n\to\infty}\int_{a}^{z_1}\phi(z,n)dz = c, \qquad for \quad a < z_1 \leqq b.$$

In the following lemmas the variable z, which is considered real, is to range as specified either over the interval $(0, z_2)$ or over the interval (z_1, z_2) where $z_1 > 0$. The complex parameter ρ is considered to range over the circles of the sequence C_n with centers at $\rho = 0$ and radii Λ_n, the argument χ being defined by the relation

$$\rho = i\Lambda_n e^{i\chi}.$$

Any arc of C_n which lies entirely on the half-plane $0 \leqq \chi \leqq \pi$, i.e., on the half-plane in which Re $(\rho) \leqq 0$, is designated by c_n. The symbol $E(z, \rho)$ is used to represent a function which is integrable and bounded for $|\rho|$ sufficiently large, and M is used to indicate a positive constant suitably chosen. Lastly, we use the symbols $O(1)$ and $o(1)$ to indicate functions which are bounded or approach zero respectively as $n \to \infty$.

LEMMA I.* *If $\psi(z)$ is any function which is integrable on the interval $(0, z_2)$ and $0 \leqq z' \leqq z'' \leqq z_2$, then*

$$\int_{z'}^{z''}\int_{c_n} e^{\rho z}E(z,\rho)\psi(z)\frac{d\rho}{\rho}dz = o(1).$$

LEMMA II. *If on the arc c_n, $\chi \leqq \chi_2$, then*

$$\int_{c_n}\{1 - e^{i\mu x}\}^{\sigma}\rho^{\tau}E(z,\rho)d\rho = \Lambda_n{}^{\tau+1}\chi_2{}^{\sigma+1}O(1).$$

This conclusion is immediately drawn upon using the fact that

(79) $$|1 - e^{i\mu x}|^{\sigma} \leqq M\chi^{\sigma}.$$

LEMMA III.† *If μ is an even integer and $z > 0$, the integral*

(80) $$I(z,n) = \int_{c_n}\{1 - e^{i\mu x}\}^{\sigma}\rho^{\tau}e^{\rho z}d\rho$$

satisfies the relation

$$I(z,n) = z^{-\sigma-1}\Lambda_n{}^{\tau-\sigma}O(1).$$

On the arc c_n we have clearly

---

* Tamarkin 2, loc. cit., p. 43.

† Stone, these Transactions, vol. 28 (1926), p. 714.

$$|e^{\rho z}| = e^{-\Lambda_n z \sin\chi} \leqq \begin{cases} e^{-(1/2)\Lambda_n z\chi} & \text{for } 0 \leqq \chi \leqq \pi/2, \\ e^{-(1/2)\Lambda_n z(\pi-\chi)} & \text{for } \pi/2 < \chi \leqq \pi. \end{cases} \tag{81}$$

Consider first the integral $I'(z, n)$ obtained by replacing c_n in (80) by that portion of c_n on which $\chi \leqq \pi/2$. We obtain then from (79) and (81) the inequality

$$|I'(z,n)| \leqq M \int_0^{\pi/2} \chi^\sigma \Lambda_n^{\tau+1} e^{-(1/2)\Lambda_n z\chi} d\chi,$$

and setting

$$\zeta = \tfrac{1}{2}\Lambda_n z\chi,$$

we find that

$$|I'(z,n)| \leqq M\left(\frac{2}{z\Lambda_n}\right)^{\sigma+1} \Lambda_n^{\tau+1} \int_0^\infty \zeta^\sigma e^{-\zeta} d\zeta = M_1 z^{-\sigma-1}\Lambda_n^{\tau-\sigma}.$$

The integral $I''(z, n)$ contributed to (80) by the integration over the portion of c_n for which $\chi > \pi/2$ may be treated similarly. Thus if we set $\pi - \chi = \eta$, then since μ is by hypothesis an even integer we have

$$e^{i\mu\chi} = e^{-i\mu\eta},$$

and the reasoning above may be repeated with $-\eta$ replacing χ. This establishes the lemma.

LEMMA IV. *If in the expression* (80) $\sigma \geqq \tau$, *then for any integrable function* $\psi(z)$ *and any expression* $[\alpha(z)]$ *of the type* (15) *with integrable coefficients*

$$\int_{z'}^{z''} I(z,n)\psi(z)[\alpha(z)]dz = o(1), \text{ for } z_1 \leqq z' \leqq z'' \leqq z_2. \tag{82}$$

From the definition (15) we derive the form

$$\psi(z)\rho^\tau[\alpha(z)] = P_\tau(z,\rho) + \frac{E(z,\rho)}{\rho}\psi(z), \tag{83}$$

where P_τ is a polynomial in ρ of degree τ. By Lemma I the last term of (83) contributes to the integral (82) a quantity of order $o(1)$. To discuss the terms of (82) due to the expression $P_\tau(z, \rho)$ we observe that Lemma III may be applied to the members on the right of the equality

$$\int_{z'}^{z''} I(z,n)dz = \int_{c_n} \{1 - e^{i\mu\chi}\}^\sigma \rho^{\tau-1} e^{\rho z''} d\rho - \int_{c_n} \{1 - e^{i\mu\chi}\}^\sigma \rho^{\tau-1} e^{\rho z'} d\rho, \tag{84}$$

to show that this integral is of order o (1), and hence that we may apply the Lebesgue theorem A to the integrals arising from the terms of the polynomial $P_\tau(z, \rho)$. The conclusion of the lemma follows.

LEMMA V. *If in the integral* (80) $\tau=0$, *and* c_n *is the semicircle* $0\leqq\chi\leqq\pi$, *then*

$$\lim_{n\to\infty}\int_0^{z_1} I(z,n)\psi(z)dz = -\pi i\psi(0+),$$

for any function $\psi(z)$ *which is integrable and is of bounded variation in some interval* $(0, \epsilon)$ *where* $\epsilon>0$.

The integration from ϵ to z_2 yields a result which is of the order $o(1)$ by Lemma IV. Hence to prove the present lemma we need only show that the integral from 0 to ϵ satisfies the hypotheses of the Lebesgue theorem B. The first of these hypotheses, (i), is clearly fulfilled, and, applying Lemma III to the integrals on the right of (84) with $\tau=0$, it is also clear that condition (ii) is fulfilled. Lastly, if we set $\phi=e^{ix}$, and recall that μ is an even integer, we obtain the formula*

$$I(z,n) = -\int_{-1}^{1}\left\{1-\phi^{\mu}\right\}^{\sigma}e^{i\Lambda_n z\phi}i\Lambda_n d\phi = -2i\int_0^1\left\{1-\phi^{\mu}\right\}^{\sigma}\Lambda_n\cos\Lambda_n z\phi d\phi,$$

from which it follows that

$$\int_0^{\epsilon} I(z,n)dz = -2i\int_0^1\left\{1-\phi^{\mu}\right\}^{\sigma}\frac{\sin\Lambda_n\epsilon\phi}{\phi}d\phi.$$

But this last integral approaches the limit $-\pi i$, as one sees by applying it to the Lebesgue theorem B itself, and using the Dirichlet integral formula

$$\lim_{n\to\infty}\int_0^1\frac{\sin\Lambda_n\epsilon\phi}{\phi}d\phi = \pi/2.$$

Hence condition (iii) is also fulfilled and the lemma is proved.

15. **The integrals** $\mathfrak{Y}_n^0(x)\cdot$ **and** $\mathfrak{Y}_n^1(x)\cdot$ **for** $0<x<1$, $\alpha=0$, **and** $\sigma_i\geqq 0$. Let the integrals $\mathfrak{Y}_n^l(x)\cdot$ be defined by the formula

$$(85)\quad \mathfrak{Y}_n^l(x)\cdot = \frac{-1}{2\pi i}\int_0^1\int_{C_n}\mathfrak{S}_n(\lambda)\,\mathcal{G}^l(x,t,\lambda)\mathfrak{R}(t)\bar{\mathfrak{F}}(t)\cdot d\lambda dt \qquad (l=0,1,2,3).$$

Then by (74) the sum of these integrals yields the integral in (65) and we may determine the limit of the latter by considering separately the evaluations of the expressions (85). The details of these evaluations differ somewhat in the cases $\alpha=0$, $\alpha=\pi/2$, and $0<\alpha<\pi/2$. They depend also upon whether x lies in the interior or at the ends of the interval $(0, 1)$. We begin by considering in detail the case in which $\alpha=0$ and $0<x<1$.

† Cf. Stone loc. cit., pp. 724-726.

If $\alpha=0$, the constants ν_1 and ν_2 in (6b) are both unity and the functions $r_1(x)$, $r_2(x)$, are accordingly real and positive. By § 13 we choose in this case $\mu=4$. Moreover, the deductions which follow require that we add to our hypotheses the condition that in the characteristic equation (35)

$$D_0(\lambda) \not\equiv 0, \qquad D_3(\lambda) \not\equiv 0 \qquad (\text{if } \alpha = 0). \tag{86}$$

From the formula for $\overline{D}(\lambda)$ given in (75a) we find the alternative forms

$$D^{-1}\overline{D}\begin{cases} = \dfrac{D_0 + D_2 e^{\lambda R_2}}{D} \\ = 1 - \dfrac{D_1 e^{\lambda R_1} + D_3 e^{\lambda R_1 + \lambda R_2}}{D}, \quad \text{for } t \leqq x, \end{cases} \tag{87}$$

while for $t>x$ these values are to be modified by subtracting 1. Now it was observed in § 7 that on the contours C_n the quotient of $D(\lambda)$ by any one of its terms is uniformly bounded from zero. Hence we have, in particular, the two evaluations

$$D^{-1}(\lambda)\begin{cases} = D_3^{-1}(\lambda) e^{-\lambda R_1 - \lambda R_2} O(1), \\ = D_0^{-1}(\lambda) O(1). \end{cases} \tag{88}$$

Let h be defined now as the largest of the numbers $\kappa_2-\kappa_3$, $\kappa_1-\kappa_0$, and $\pm(\kappa_3-\kappa_0)$, where for any l the number $\kappa_j-\kappa_l$ is to be considered zero if $D_j(\lambda)\equiv 0$. Then substituting the values (88) in the respective formulas (87), supposing $\mathrm{Re}(\lambda)\leqq 0$ in the first formula and $\mathrm{Re}(\lambda)>0$ in the second, we find that

$$D^{-1}\overline{D}\begin{cases} = \lambda^h e^{-\lambda R_1} O(1), & \text{for } \mathrm{Re}\,(\lambda) > 0, \\ = 1 - \lambda^h e^{\lambda R_1} O(1), & \text{for } \mathrm{Re}\,(\lambda) \leqq 0, \quad t \leqq x, \end{cases} \tag{89}$$

the values to be diminished by one when $t>x$.

With the value of $\mathcal{G}^0(x, t, \lambda)$ obtained from (75a) with the use of (89) and (76) we may now write the integral (85) for $l=0$ in the explicit form

$$\begin{aligned} \mathcal{Y}_n^0(x) \cdot = \frac{-1}{2\pi i}\Big(& \int_0^x \int_{C_n'} s_{in}(\lambda)\lambda^{-\delta_{i2}}\omega_{i,11}(x,t,\lambda) e^{\lambda R_1(x,t)} d\lambda dt \\ & - \int_x^1 \int_{C_n''} s_{in}(\lambda)\lambda^{-\delta_{i2}}\omega_{i,11}(x,t,\lambda) e^{\lambda R_1(x,t)} d\lambda dt \\ & - \int_0^1 \int_{C_n'} s_{in}\lambda^{h-\delta_{i2}}\omega_{i,11} e^{\lambda R_1(x)+\lambda R_1(1,t)} O(1) d\lambda dt \\ & + \int_0^1 \int_{C_n''} s_{in}\lambda^{h-\delta_{i2}}\omega_{i,11} e^{-\lambda R_1(1,x)-\lambda R_1(t)} O(1) d\lambda dt \Big) \cdot, \end{aligned} \tag{90}$$

where C_n' is the part of the contour C_n on which $\mathrm{Re}(\lambda) \leqq 0$ and C_n'' is the part on which $\mathrm{Re}(\lambda) > 0$.

Consider first the final integral on the right of this equation (90). The quantity $\{R_1(1, x) + R_1(t)\}$ is positive for all values of t on the interval $(0, 1)$. Hence it may assume the rôle of the variable z in Lemma IV, and if we substitute for $\omega_{i,11}(x, t, \lambda)$ its value as given by (77) it becomes obvious that the lemma in question applies, and hence that the integral is of the order $o(1)$ provided $\sigma_i \geqq h - \delta_{i2}$. A similar argument with the same condition on σ_i shows that the integral preceding that just considered is likewise of the order $o(1)$.

Consider now the leading integral on the right of (90). In the case of the component $i = 2$ the integral is of the type discussed by Lemma I, the rôle of the variable z being taken by the quantity $R_1(x, t)$, and it follows, therefore, by Lemma I that the integral is of order $o(1)$. This argument, moreover, may be repeated for the component $i = 1$ provided the first term of the expression $\omega_{1,11}(x, t, \lambda)$ as given by (77) is tentatively omitted from the consideration. The discussion of the second integral in (90) is entirely analogous to that of the first, and hence recalling the omitted first term of $\omega_{1,11}$ we find that formula (90) reduces to the form

$$\mathfrak{F}_n^0(x) \cdot = \frac{-1}{2\pi i}\Big(\delta_{i1}\int_0^x \int_{C_n'} s_{1n}(\lambda)\bar{f}_1(t)e^{\lambda R_1(x,t)}r_1(t)d\lambda dt - \delta_{i1}\int_x^1 \int_{C_n''} s_{1n}(\lambda)\bar{f}_1(t)e^{\lambda R_1(x,t)}r_1(t)d\lambda dt + o(1)\Big)\cdot . \tag{91}$$

To reveal more clearly the structure of the integrals in this formula we resort to the change of variable defined by the relations

$$R_1(t) = \xi_0 - \xi, \quad R_1(x) = \xi_0, \quad R_1 = \xi_0 - \xi, \quad \bar{f}_1(t) = \psi_1(\xi_0 - \xi). \tag{92}$$

Substituting these values and subsequently replacing ξ and λ in the second integral by $-\xi$ and $-\lambda$, respectively, we find for (91) the form

$$\mathfrak{F}_n^0(x) \cdot = \frac{-1}{2\pi i}\Big(\delta_{i1}\int_0^{\xi_0} \int_{C_n'} s_{1n}(\lambda)e^{\lambda\xi}\psi_1(\xi_0 - \xi)d\lambda d\xi + \delta_{i1}\int_0^{\xi_1} \int_{C_n'} s_{1n}(\lambda)e^{\lambda\xi}\psi_1(\xi_0 + \xi)d\lambda d\xi + o(1)\Big)\cdot . \tag{93}$$

Let us suppose now that x is a point in some neighborhood of which $\bar{f}_1(t)$ is of bounded variation. Then it is clear from (92) that $\psi_1(\xi_0 \pm \xi)$ is of bounded variation in some neighborhood of $\xi = 0$, and by Lemma V the

integrals in (93) converge respectively to the values $-\pi i\psi_1(\xi_0 \pm \xi)$. Since $\psi_1(\xi_0 \pm \xi) \equiv \bar{f}_1(x \pm 0)$, we have the result

$$\mathcal{J}_n^0(x)\cdot \to \frac{1}{2}\begin{pmatrix} \bar{f}_1(x-) + \bar{f}_1(x+) \\ 0 \end{pmatrix}\cdot . \tag{94}$$

The discussion thus given for the integral (85) with $l=0$ may be repeated almost without change for the integral with $l=1$. We find from such a discussion that if x is a point in some neighborhood of which $\bar{f}_2(t)$ is of bounded variation, and if $\sigma_i \geqq h_1 - \delta_{i1}$, where h_1 designates the largest of the numbers $\kappa_1 - \kappa_3$, $\kappa_2 - \kappa_0$, $\pm(\kappa_3 - \kappa_0)$, then

$$\mathcal{J}_n^1(x)\cdot \to \frac{1}{2}\begin{pmatrix} 0 \\ \bar{f}_2(x-) + \bar{f}_2(x+) \end{pmatrix}\cdot . \tag{95}$$

Combining these results and recalling the relation between the vectors $\overline{\mathfrak{F}}(x)\cdot$ and $\mathfrak{F}(x)\cdot$, as given by (57), we obtain finally the result

$$\mathcal{A}(x)\cdot + \overline{\mathcal{J}}_n^0(x)\cdot + \mathcal{J}_n^1(x)\cdot \to \tfrac{1}{2}\,\mathfrak{F}(x+)\cdot + \tfrac{1}{2}\,\mathfrak{F}(x-)\cdot .$$

16. **The integrals $\mathcal{J}_n^2(x)\cdot$ and $\mathcal{J}_n^3(x)\cdot$ for $0<x<1$, $\alpha=0$, and $\sigma_i \geqq 0$.** The integrals (85) for $l=2, 3$ are now easily evaluated by the method of the preceding section. Thus we have

$$D^{-1}D_4 \begin{cases} = \lambda^{\kappa_4-\kappa_3} e^{-\lambda R_1 - \lambda R_2} O(1), & \text{for } \operatorname{Re}(\lambda) > 0, \\ = \lambda^{\kappa_4-\kappa_0} O(1), & \text{for } \operatorname{Re}(\lambda) \leqq 0, \end{cases}$$

and we find in the manner now familiar that

$$\mathcal{J}_n^2(x)\cdot = \frac{-1}{2\pi i}\Big(\int_0^1 \int_{C_n'} s_{in}(\lambda)\lambda^{\kappa_4-\kappa_0-\delta_{i2}}\omega_{i,12}(x,t,\lambda) e^{\lambda R_1(x)+\lambda R_2(1,t)} O(1)\,d\lambda\,dt$$
$$+ \int_0^1 \int_{C_n''} s_{in}(\lambda)\lambda^{\kappa_4-\kappa_3-\delta_{i2}}\omega_{i,12}(x,t,\lambda) e^{-\lambda R_1(1,x)-\lambda R_2(t)} O(1)\,d\lambda\,dt\Big)\cdot .$$

An application of Lemma IV yields the result that these integrals are both of the order $o(1)$ provided $\sigma_i \geqq \kappa_4 - \kappa_0 - \delta_{i2}$, $\kappa_4 - \kappa_3 - \delta_{i2}$. In entirely similar fashion it is found that

$$\mathcal{J}_n^3(x)\cdot \to \mathcal{O},$$

provided $\sigma_i \geqq \kappa_5 - \kappa_0 - \delta_{i1}$, $\kappa_5 - \kappa_3 - \delta_{i1}$, and the results of this and the preceding section may, therefore, be summarized as follows:

THEOREM 1. *If $\alpha=0$ and $D_0(\lambda)D_3(\lambda) \not\equiv 0$, let $\bar{\kappa}_i$, $i=1, 2$, designate the largest of the numbers*

$$\pm(\kappa_3 - \kappa_0), \quad \kappa_2 - \kappa_3 - \delta_{i2}, \quad \kappa_1 - \kappa_0 - \delta_{i2}, \quad \kappa_4 - \kappa_3 - \delta_{i2}, \quad \kappa_4 - \kappa_0 - \delta_{i2},$$
$$\kappa_1 - \kappa_3 - \delta_{i1}, \quad \kappa_2 - \kappa_0 - \delta_{i1}, \quad \kappa_5 - \kappa_3 - \delta_{i1}, \quad \kappa_5 - \kappa_0 - \delta_{i1},$$

the number $\kappa_j - \kappa_l$ *being considered zero for any* l *if* $D_j(\lambda) \equiv 0$. *Then with* $\mu = 4$ *and* σ_i *chosen in any way subject to the relation* $\sigma_i \geqq \bar{\kappa}_i, i = 1, 2$, *the expansion* (60) *associated with an arbitrary integrable vector* $\mathfrak{F}(x)\cdot$ *is summable of order* (σ_1, σ_2) *to the value*

$$\frac{1}{2}\{\mathfrak{F}(x+)\cdot + \mathfrak{F}(x-)\cdot\}, \tag{96}$$

at every point x *within the interval* $(0, 1)$, *in some neighborhood of which* $\mathfrak{F}(x)\cdot$ *is of bounded variation.*

COROLLARY 1. *If* $\bar{\kappa}_i = 0, i = 1, 2$, *the expansion* (60) *is convergent.*

COROLLARY 2. *For an arbitrary choice of the integrable functions* $f_1(x)$, $f_2(x)$, *the expansion* $I_{i,n}(f_1, f_2)$ *given by* (72) *for either* $i = 1$, *or* $i = 2$, *is summable of order* $\sigma_i \geqq \bar{\kappa}_i$ *to the value*

$$\frac{1}{2}\{f_i(x+) + f_i(x-)\},$$

at any point within the interval $(0, 1)$, *in some neighborhood of which* $f_i(x)$ *is of bounded variation.*

In the following section we shall establish the summability of the expansion (in contradistinction to its actual convergence) without hypothesizing the boundedness of the variation of $\mathfrak{F}(x)\cdot$. In this respect, therefore, the statement of Theorem 1 may be broadened.

17. **The expansion (60) when $0 < x < 1$, $\alpha = 0$ and $\sigma_i > 0$.** It follows from Lemma IV, since the function $\psi_1(\xi_0 \pm \xi)$ of §15 is integrable, that with an arbitrary choice of a positive ϵ, and $\xi > \epsilon$,

$$\int_\epsilon^\xi \int_{C_{n'}} s_{1n}(\lambda) e^{\lambda\xi} \psi_1(\xi_0 \pm \xi) d\lambda d\xi = o(1).$$

Hence we may also write the formula (93) in the form

$$\mathfrak{Y}_n^0(x) = \frac{-1}{2\pi i}\left(\delta_{i1} \int_0^\epsilon I(\xi, n)\{\psi_1(\xi_0 - \xi) + \psi_1(\xi_0 + \xi)\} d\xi + o(1)\right)\cdot, \tag{97}$$

where we have set

$$I(\xi, n) = \int_{C_{n'}} s_{1n}(\lambda) e^{\lambda\xi} d\xi.$$

This integral $I(\xi, n)$ is of the form considered in Lemmas II and III with $\tau = 0$, and we draw from these lemmas the inequalities

$$\begin{aligned} &|I(\xi, n)| < M\Lambda_n, \\ &|I(\xi, n)| < M\Lambda_n^{-\sigma_1}\xi^{-\sigma_1 - 1}, \quad \text{for} \quad \xi > 0, \end{aligned} \tag{98}$$

M being a suitable positive constant.

Let $\Psi_1(\xi_0,\xi)$ and $\Phi_1(\xi_0, \xi)$ be defined now by the relations

$$\Psi_1(\xi_0,\xi) = \psi_1(\xi_0 - \xi) + \psi_1(\xi_0 + \xi) - 2\psi(\xi_0),$$

$$\Phi_1(\xi_0,\xi) = \frac{1}{\xi}\int_0^{\xi} |\Psi_1(\xi_0,\xi)|\, d\xi.$$

Then by a theorem of Lebesgue* the function $\Phi_1(\xi_0, \xi)$ approaches zero with ξ for almost all values of ξ_0. We except from consideration the values of ξ_0, if any, for which this is not true. Then with an arbitrary choice of an $\eta>0$ the ϵ in formula (97) may be chosen so that

$$\Phi_1(\xi_0,\xi) < \eta, \quad \text{for} \quad 0 \leqq \xi \leqq \epsilon. \tag{99}$$

Let the formula (97) be written now in the form

$$\mathfrak{F}_n^0(x)\cdot = \frac{-1}{2\pi i}\left(2\psi(\xi_0)\delta_{i1}\int_0^{*} I(\xi,n)d\xi + \delta_{i1}\int_0^{*} I(\xi,n)\Psi_1(\xi_0,\xi)d\xi + o(1)\right)\cdot. \tag{100}$$

From Lemma V we find directly for the first integral in this formula the value

$$\int_0^{*} I(\xi,n)d\xi = -\pi i + o(1).$$

The second integral, on the other hand, we evaluate by the following method which is due to Hardy.†

From the first of formulas (98), together with (99), we find that for n sufficiently large

$$\left|\int_0^{1/\Lambda_n} I(\xi,n)\Psi_1(\xi_0,\xi)d\xi\right| < M\Lambda_n\int_0^{1/\Lambda_n} |\Psi_1(\xi_0,\xi)|\, d\xi = M\Phi_1(\xi_0,1/\Lambda_n) < M\eta.$$

Moreover, an application of the second of formulas (98) followed by an integration by parts yields, when n is sufficiently large and $\sigma_1>0$, the inequality

$$\begin{aligned}\left|\int_{1/\Lambda_n}^{*} I(\xi,n)\Psi(\xi_0,\xi)d\xi\right| &< M\Lambda_n^{-\sigma_1}\int_{1/\Lambda_n}^{*}\frac{|\Psi_1(\xi_0,\xi)|}{\xi^{\sigma_1+1}}d\xi \\ &= M\Lambda_n^{-\sigma_1}\left\{\frac{\Phi_1(\xi_0,\xi)}{\xi^{\sigma_1}}\Big|_{1/\Lambda_n}^{*} + (\sigma_1+1)\int_{1/\Lambda_n}^{*}\frac{\Phi_1(\xi_0,\xi)}{\xi^{\sigma_1+1}}d\xi\right\}, \\ &< M\left\{\frac{\Phi_1(\xi_0,\xi)}{(\Lambda_n\epsilon)^{\sigma_1}} + \Phi_1(\xi_0,1/\Lambda_n) + (\sigma_1+1)\eta\Lambda_n^{-\sigma}\int_{1/\Lambda_n}^{*}\frac{d\xi}{\xi^{\sigma_1+1}}\right\} < M_1\eta.\end{aligned}$$

* Lebesgue, *Leçons sur les Séries Trigonométriques*, Paris, 1905, p. 96.

† Hardy, Proceedings of the London Mathematical Society, (2), vol. 12 (1913), p. 368.

It follows that when σ_1 is positive the second integral in (100) is of the order $o(1)$, and hence that at almost all points within the interval

$$\mathfrak{Y}_n^0(x)\cdot \to \begin{pmatrix} \bar{f}_1(x) \\ 0 \end{pmatrix}\cdot .$$

The discussion of $\mathfrak{Y}_n^1(x)\cdot$ is precisely similar to that of $\mathfrak{Y}_n^0(x)\cdot$ just completed. Since the argument of §16 was based only on the integrability of the functions $f_1(x), f_2(x)$, the reasoning of that section requires no modification and we may summarize the results as follows.

THEOREM 2. *If the conditions of Theorem* 1 *are satisfied, and if, furthermore,* $\sigma_i>0$, $i=1, 2$, *then the expansion* (60) *is summable of order* (σ_1, σ_2) *to the value* $\mathfrak{F}(x)\cdot$ *at almost all points* x *in the interior of the interval* (0, 1).

COROLLARY. *If the conditions of Theorem* 1, *Corollary* 2, *are satisfied and if, furthermore,* $\sigma_i>0$, *then the expansion* $I_{in}(f_1, f_2)$, $i=1, 2$, *is summable of order* σ_i *to the value* $f_i(x)$ *at almost all points within the interval* (0, 1).

18. **The expansion when $\alpha=\pi/2$, and $0<x<1$.** When $\alpha=\pi/2$, the rays (37) coincide with the axis of reals, and the constants ν_i of formula (6b) have the values $\nu_1=i$, $\nu_2=-i$. If it is observed that in this case

$$D(\lambda)e^{\lambda i|R_2|}$$

is similar in structure to $D(\lambda)$ in the case $\alpha=0$, with λ replaced by λi and R_j replaced by $|R_j|$, it becomes evident that the discussions of §§15, 16, and 17 are easily adapted to apply to the case in hand. The hypothesis to be made in correspondence to (86) is that

$$D_1(\lambda) \not\equiv 0, \qquad D_2(\lambda) \not\equiv 0 \qquad (\text{if } \alpha = \pi/2),$$

and by §13 we are to choose $\mu=2$. The resulting theorems are the following.

THEOREM 3. *If* $\alpha=\pi/2$ *and* $D_1(\lambda)D_2(\lambda)\not\equiv 0$, *let* $\bar{\kappa}_i$, $i=1, 2$, *designate the largest of the numbers*

$$\pm(\kappa_2-\kappa_1),\quad \kappa_3-\kappa_2-\delta_{i2},\quad \kappa_0-\kappa_1-\delta_{i2},\quad \kappa_4-\kappa_2-\delta_{i2},\quad \kappa_4-\kappa_1-\delta_{i2},$$
$$\kappa_0-\kappa_2-\delta_{i1},\quad \kappa_3-\kappa_1-\delta_{i1},\quad \kappa_5-\kappa_2-\delta_{i1},\quad \kappa_5-\kappa_1-\delta_{i1}.$$

Then with $\mu=2$ *and* $\sigma_i\geqq\bar{\kappa}_i$ *the expansion* (60) *is summable of order* (σ_1, σ_2) *to the value* (96) *at every point* x *within the interval* (0, 1) *in some neighborhood of which* $\mathfrak{F}(x)\cdot$ *is of bounded variation.*

THEOREM 4. *If the conditions of Theorem* 3 *are satisfied and if, furthermore,* $\sigma_i>0$, $i=1, 2$, *then the expansion is summable of order* (σ_1, σ_2) *to the value* $\mathfrak{F}(x)\cdot$ *at almost all points within the interval* (0, 1).

It is clear that corollaries similar to those of Theorems 1 and 2, may also be appended respectively to the Theorems 3 and 4.

19. **The expansion when $0<\alpha<\pi/2$, and $0<x<1$.** The details of the discussion of the integrals (85) differ somewhat when $0<\alpha<\pi/2$ from those of the preceding sections. The hypothesis to be made for this case is that no term of the characteristic equation (35) is lacking, i.e.,

$$D_j(\lambda) \not\equiv 0, \quad j = 0,1,2,3 \quad (\text{if } 0 < \alpha < \pi/2).$$

Moreover, as observed in §13, we shall, in the case of summability as distinguished from that of convergence of the expansion, restrict the consideration by the further hypothesis that α is commensurable with π.

The rays (37) together with the axes of reals and imaginaries divide the contour C_n into eight arcs which we designate by $C_n^{(j,i)}$, where $C_n^{(j,1)}$ and $C_n^{(j,2)}$ are the arcs on the jth quadrant taken in counter-clockwise succession. Since $R_j(x)=\nu_j|R_j(x)|$, whereas, on the arc $C_n^{(1,1)}$, we have $\mathrm{Re}\ (\lambda\nu_1)\leqq 0$, $\mathrm{Re}(\lambda\nu_2)>0$, we may write

$$D^{-1}(\lambda) = D_3^{-1}(\lambda)e^{-\lambda\nu_1|R_1|-\lambda\nu_2|R_2|}O(1), \quad \text{on arc } C_n^{(1,1)}.$$

On the arc $C_n^{(1,2)}$ we have, on the other hand, $\mathrm{Re}\ (\lambda\nu_1)\leqq 0$, $\mathrm{Re}\ (\lambda\nu_2)>0$, and we write accordingly

$$D^{-1}(\lambda) = D_2^{-1}(\lambda)e^{-\lambda\nu_2|R_2|}O(1), \quad \text{on arc } C_n^{(1,2)}.$$

With these values substituted respectively into the formulas (87), we obtain the relations

$$D^{-1}\overline{D}\begin{cases} = \lambda^{\kappa_2-\kappa_3}e^{-\lambda\nu_1|R_1|}O(1), & \text{on arc } C_n^{(1,1)}, \\ = 1-\lambda^{\kappa_2-\kappa_1}e^{\lambda\nu_1|R_1|}O(1), & \text{on arc } C_n^{(1,2)}, \quad \text{for} \quad t\leqq x, \end{cases}$$

the values to be diminished by one for $t>x$.

On the portion of the contour C_n in the second quadrant, we use the second of equations (87) together with the evaluations

$$D^{-1}(\lambda) = D_0^{-1}(\lambda)O(1), \quad \text{when} \quad D_1(\lambda) \geqq D_3(\lambda)e^{\lambda\nu_2|R_2|},$$

and

$$D^{-1}(\lambda) = D_2^{-1}(\lambda)e^{-\lambda\nu_2|R_2|}O(1), \quad \text{when} \quad D_1(\lambda) < D_3(\lambda)e^{\lambda\nu_2|R_2|},$$

and obtain in this way the formula

(101) $D^{-1}\overline{D} = 1-\lambda^{h_0}e^{\lambda\nu_1|R_1|}O(1)$, on arcs $C_n^{(2,1)}$ and $C_n^{(2,2)}$, for $t\leqq x$.

Here h_0 designates the larger of the numbers $\kappa_3-\kappa_2$ and $\kappa_1-\kappa_0$. In similar fashion we find also the further formulas

$$(102)\quad D^{-1}\overline{D}\begin{cases} = 1 - \lambda^{\kappa_3-\kappa_2}e^{\lambda\nu_1|R_1|}O(1), & \text{on arc } C_n^{(3,1)} \\ = \lambda^{\kappa_3-\kappa_1}e^{-\lambda\nu_1|R_1|}O(1), & \text{on arc } C_n^{(3,2)} \\ = \lambda^{h_1}e^{-\lambda\nu_1|R_1|}O(1), & \text{on arcs } C_n^{(4,1)} \text{ and } C_n^{(4,2)}, \text{ for } t \leqq x, \end{cases}$$

with h_1 defined as the larger of the numbers $\kappa_0-\kappa_1$ and $\kappa_2-\kappa_3$. In both formulas (101) and (102) the values given are to be diminished by one when $t>x$.

The integral (85) with $l=0$ may now be written explicitly, and is found to be given by the formula

$$\begin{aligned}(103)\quad \mathfrak{F}_n^0(x)\cdot = \frac{-1}{2\pi i}\Big(&\int_0^x \int_{C_n'} s_{in}\lambda^{-\delta_{i2}}\omega_{i,11}(x,t,\lambda)e^{\lambda\nu_1|R_1(x,t)|}d\lambda dt \\ &- \int_x^1 \int_{C_n''} s_{in}\lambda^{-\delta_{i2}}\omega_{i,11}(x,t,\lambda)e^{\lambda\nu_1|R_1(x,t)|}d\lambda dt \\ &+ \int_0^1 \int_{C_n'} s_{in}\lambda^{h-\delta_{i2}}\omega_{i,11}e^{\lambda\nu_1|R_1(x)|+\lambda\nu_1|R_1(1,t)|}O(1)d\lambda dt \\ &+ \int_0^1 \int_{C_n''} s_{in}\lambda^{h-\delta_{i2}}\omega_{i,11}e^{-\lambda\nu_1|R_1(1,x)|-\lambda\nu_1|R_1(t)|}O(1)d\lambda dt\Big)\cdot, \end{aligned}$$

h being the largest of the numbers $\pm(\kappa_3-\kappa_2)$ and $\pm(\kappa_1-\kappa_0)$, while C_n' and C_n'' represent respectively the portions of C_n on which $\text{Re}\,(\lambda\nu_1)\leqq 0$ and on which $\text{Re}\,(\lambda\nu_1)>0$. We observe now that this formula (103) differs from (90) only in that λ is replaced by $\lambda\nu_1$ and R_j by $|R_j|$. Hence the discussion given subsequent to formula (90) may be repeated with slight modifications as applied to the formula (103) and as conclusion we obtain again the relation (94), provided in this case that $\sigma_i \geqq \pm(\kappa_3-\kappa_2)-\delta_{i2}$, $\pm(\kappa_1-\kappa_0)-\delta_{i2}$.

In precisely similar fashion the result (95) may now be derived under the condition that $\sigma_i \geqq \pm(\kappa_2-\kappa_0)-\delta_{i1}$, $\pm(\kappa_3-\kappa_1)-\delta_{i1}$, and the conclusion that $\mathfrak{F}_n^2(x)\cdot$ and $\mathfrak{F}_n^3(x)\cdot$ each approach $\mathcal{O}$ as a limit follows if $\sigma_i \geqq \kappa_4-\kappa_j-\delta_{i2}$, $\kappa_5-\kappa_j-\delta_{i1}$, $j=0, 1, 2, 3$. The results, then, may be summarized as follows.

THEOREM 5. *If* $0<\alpha<\pi/2$ *and* $D_j(\lambda)\not\equiv 0$, $j=0, 1, 2, 3$, *let* $\bar{\kappa}_i$, $i=1, 2$, *designate the largest of the numbers*

$$\begin{aligned} &\pm\ (\kappa_3-\kappa_2)-\delta_{i2}, \qquad \pm\ (\kappa_1-\kappa_0)-\delta_{i2}, \qquad \kappa_4-\kappa_j-\delta_{i2}, \\ &\pm\ (\kappa_2-\kappa_0)-\delta_{i1}, \qquad \pm\ (\kappa_3-\kappa_1)-\delta_{i1}, \qquad \kappa_5-\kappa_j-\delta_{i1}, \quad j=0,1,2,3. \end{aligned}$$

Then (i) *if* $\bar{\kappa}_i=0$ *the expansion converges to the value* (96) *at every point in the interior of the interval* (0, 1) *in some neighborhood of which* $\mathfrak{F}(x)\cdot$ *is of bounded variation*; (ii) *if* $\bar{\kappa}>0$, *and furthermore* α/π *is rational, then with* $\sigma_i \geqq \bar{\kappa}_i$ *the expansion is summable of order* (σ_1, σ_2) *to the value* (96) *at every inner point of* (0, 1) *in some neighborhood of which* $\mathfrak{F}(x)\cdot$ *is of bounded variation.*

THEOREM 6. *If the conditions of Theorem* 5 *are satisfied and* $\sigma_i > 0$, *the expansion is summable of order* (σ_1, σ_2) *to the value* $\mathfrak{F}(x)\cdot$ *at almost all points within the interval* (0, 1).

The analogous facts for the expansions (72) may obviously be covered in corollaries to these theorems as was done in the case of Theorems 1 and 2.

20. **The expansion at $x=1$ when $\alpha=0$.** The argument from which the convergence of the integrals in the preceding sections was concluded depends explicitly upon the fact that x is an interior point of the interval (0, 1). Not only must the reasoning be elaborated when x is an end point of the interval, but we shall be able to conclude summability of the expansion only in the case of functions $f_1(x)$, and $f_2(x)$, satisfying suitable restrictions. For the sake of brevity we shall confine our considerations to the case $x=1$ and $\alpha=0$. The discussions required for the case $x=0$, and for other values of α may be given in an entirely analogous manner. To simplify the statement of the results we shall consider $\mathfrak{F}(x)\cdot$ defined at $x=0$ and $x=1$ so that $\mathfrak{F}(0)\cdot = \mathfrak{F}(0+)\cdot$, and $\mathfrak{F}(1)\cdot = \mathfrak{F}(1-)\cdot$ whenever these limits exist.

We suppose, then, that the conditions of Theorem 1 are fulfilled. Then a review of the integrals in the expressions for $\mathfrak{F}_n'(x)\cdot$ in §§15 and 16 reveals the fact that the integrals over the arc C_n' are of the same character for $x=1$ as for $0<x<1$, and hence that they converge to the value $\frac{1}{2}\mathfrak{F}(1)\cdot$. But from formula (58) this value is zero, and hence we need consider these integrals no further. The integrals over the arc C_n'', however, do require further consideration. Thus the final integral in formula (90) was evaluated by means of Lemma IV, while the applicability of that lemma depends upon the fact that $\{R_1(1, x)+R_1(t)\}>0$, a condition which is not fulfilled when $x=1$. The discussion of the integral by other means requires a more precise expression for the integrand. To obtain this we shall make use of the identity

$$\frac{1}{D(\lambda)} = \frac{1}{D_3(\lambda)e^{\lambda R_1+\lambda R_2}}\left\{-\frac{D_0 + D_1e^{\lambda R_1} + D_2e^{\lambda R_2}}{D} + 1\right\}. \tag{104}$$

Consider the fraction within the brace of this expression. It is obviously bounded. We may, however, also obtain an evaluation of it by utilizing the first of formulas (88), and hence (104) may be written in the form

$$D^{-1}(\lambda) = \lambda^{-\alpha_3}e^{-\lambda R_1-\lambda R_2}N(\lambda) + D_3^{-1}e^{-\lambda R_1-\lambda R_2}, \tag{105}$$

in which $N(\lambda)$ is subject to the alternative evaluations

$$N(\lambda)\begin{cases} = O(1), \\ = \lambda^{\alpha_3-\alpha_3}e^{-\lambda R_1-\lambda R_2}O(1) + \lambda^{\alpha_1-\alpha_3}e^{-\lambda R_2}O(1) + \lambda^{\alpha_3-\alpha_3}e^{-\lambda R_1}O(1). \end{cases} \tag{106}$$

We return now to the integrals over arc C_n'' in formula (90). When $x=1$ the first of these integrals clearly does not appear. On the other hand, we obtain in place of the second integral when the first of formulas (88) and the formula (105) are substituted in the respective terms of $D^{-1}\overline{D}$ the three integrals

$$\int_0^1 \int_{C_n''} s_{in}\lambda^{\kappa_2-\kappa_3-\delta_{i2}} e^{-\lambda R_2-\lambda R_1(t)} O(1)\omega_{i,11}(1,t,\lambda)d\lambda dt, \tag{107}$$

$$\int_0^1 \int_{C_n''} s_{in}\lambda^{\kappa_2-\kappa_3-\delta_{i2}} e^{-\lambda R_1(t)} [\alpha_2] N(\lambda)\omega_{i,11}(1,t,\lambda)d\lambda dt, \tag{108}$$

$$-\int_0^1 \int_{C_n''} s_{in}\lambda^{\kappa_2-\kappa_3-\delta_{i2}} e^{-\lambda R_1(t)} \left[\frac{\alpha_2}{\alpha_3}\right] \omega_{i,11}(1,t,\lambda)d\lambda dt. \tag{109}$$

The first of these, i.e., (107), is found directly to be of order $o(1)$ by the arguments of §15. The second, however, i.e., (108), requires a more careful analysis. We shall discuss it in detail only in so far as the integration over the part of C_n'' which lies in the first quadrant is concerned. No essential modification of the argument is necessary to adapt it to the discussion of the integration over the remaining part of C_n''. To begin with we set

$$\lambda = i\Lambda_n e^{-ix}.$$

Then the arc of C_n'' on the first quadrant is given by $0 \leqq \chi \leqq \pi/2$, and this we divide by a point λ_a into arc (1) and arc (2) as follows:

arc (1): $0 \leqq \chi \leqq \chi_a$,

arc (2): $\chi_a < \chi \leqq \pi/2$,

where

$$\chi_a = \frac{A \log \Lambda_n}{\Lambda_n}.$$

In this we tentatively specify A merely as a positive constant.

On the arc (1) now we use for $N(\lambda)$ the first of the evaluations (106). By Lemma II it follows then that the integration over this arc contributes to (108) a quantity which is of the order

$$\Lambda_n^{\kappa_2-\kappa_3-\delta_{i2}} \left\{\frac{A \log \Lambda_n}{\Lambda_n}\right\}^{\sigma_i} O(1),$$

i.e., of the order $o(1)$ for any choice of the constant A provided $\sigma_i > \kappa_2-\kappa_3-\delta_{i2}$. Since the condition $\sigma_i \geqq \bar{\kappa}_i$ imposed in Theorem 1 admits also of the possibility $\sigma_i = \kappa_2-\kappa_3-\delta_{i2}$, we must also consider this case. However, the argument requires modification only in the case of the integral arising from the first term of $\omega_{i,11}(1, t, \lambda)$, and we may conclude this integral also to be of the order

$o(1)$ by the use of the Lebesgue theorem B if we suppose the function $f_1(t)$ to be of bounded variation in some neighborhood of $t=0$.

We turn next to the contribution to (108) by the integration over the arc (2). To begin with we observe that on this arc

$$\left| \lambda^{AR_j} e^{-\lambda R_j} \right| = \left| e^{(\lambda_a-\lambda)R_j} \right| \cdot \left| \lambda^{AR_j} e^{-\lambda_a R_j} \right|,$$

while in this expression the first factor on the right is bounded since $\mathrm{Re}(\lambda_a-\lambda)\leqq 0$. The second factor, on the other hand, may be written in the form

$$e^{\Lambda_n R_j(\chi_a-\sin\chi_a)},$$

from which it is seen to approach unity as $\Lambda_n\to\infty$. It follows, therefore, in virtue of the second of formulas (106), that if A is chosen sufficiently large we have on arc (2)

$$\lambda^{\kappa_2-\kappa_3-\delta_{i2}} e^{-\lambda R_1(t)}[\alpha_2]N(t) = \lambda^{-1}O(1).$$

By Lemma I then the integration over this arc contributes to (108) a quantity which is of the order $o(1)$, and in consequence the entire integral (108) is of that order.

The integral (109) remains to be evaluated to complete the discussion of $\mathcal{J}_n^0(1)\cdot$. For this evaluation and that of the corresponding integrals in the expressions $\mathcal{J}_n^l(1)\cdot$, $l=1, 2, 3$, we shall impose upon the functions $f_1(t)$ and $f_2(t)$ the restrictions that the limits

$$\lim_{t\to 0}\frac{\bar{f}_1(t)}{t^{\tau_{i1}}}, \quad \text{and} \quad \lim_{t\to 0}\frac{\bar{f}_2(t)}{t^{\tau_{i2}}}, \quad \text{exist for } i=1,2, \tag{110}$$

where τ_{ij} designates the largest of the numbers

$$\kappa_2-\kappa_3-\delta_{i2}-\delta_{j2},\ \kappa_1-\kappa_3-\delta_{i1}-\delta_{j1},\ \kappa_4-\kappa_3-\delta_{i2}-\delta_{j1},\ \kappa_5-\kappa_3-\delta_{i1}-\delta_{j2}.$$

A review of the hypothesis of Theorem 1 shows that the numbers τ_{ij} do not exceed the value of $\bar{\kappa}_i$ and hence that any σ_i chosen in accordance with Theorem 1 satisfies the relation $\sigma_i\geqq\tau_{ij}$. We shall show that if the hypothesis (110) is satisfied then the expansion is summable at $x=1$.

Consider then the integral (109). Using formula (77) we may write

$$\begin{aligned}\lambda^{\kappa_2-\kappa_3-\delta_{i2}}\left[\frac{\alpha_2}{\alpha_3}\right]\omega_{i,11}(1,t,\lambda)&\\ = \bar{f}_1(t)\sum_{m=0}^{\kappa_2-\kappa_3-\delta_{i2}}\lambda^m\beta_{im}^{(1)}(t) + \bar{f}_2(t)\sum_{m=0}^{\kappa_2-\kappa_3-\delta_{i2}-1}\lambda^m\beta_{im}^{(2)}(t)&\\ + \frac{E(t,\lambda)\bar{f}_1(t) + E(t,\lambda)\bar{f}_2(t)}{\lambda},&\end{aligned} \tag{111}$$

the coefficients $\beta_{ij}^{(k)}(t)$ as well as the functions $E(t, \lambda)$ being integrable on the interval $(0, 1)$. The integrals arising in (109) from the final terms on the right of (111) are of order $o(1)$ by Lemma I. Of the integrals due to the remaining terms in (111) we consider next that arising from the general term involving $\bar{f}_1(t)$, i.e.,

$$\int_0^1 \int_{C_n''} s_{in}(\lambda)\lambda^m e^{-\lambda R_1(t)}\beta_{im}^{(1)}(t)\bar{f}_1(t)d\lambda dt. \tag{112}$$

For any integer l not exceeding σ_i we readily verify that

$$\frac{d^l}{d\lambda^l}\{s_{in}(\lambda)\lambda^m\} = \left\{1 - \left(\frac{\lambda}{\Lambda_n}\right)^\mu\right\}^{\sigma_i - l}\lambda^{m-l}P\left(\frac{\lambda^\mu}{\Lambda^\mu}\right),$$

where P is used to designate a polynomial. Since the right-hand member of this equation vanishes at the ends of the arc C_n' when $\sigma_i > l$, we obtain for (112) by an integration by parts τ_{i1} times repeated the form

$$\int_0^1 \int_{C_n''} \left\{1 - \left(\frac{\lambda}{\Lambda_n}\right)^\mu\right\}^{\sigma_i - \tau_{i1}} \lambda^{m-\tau_{i1}} e^{-\lambda R_1(t)}\, P\left(\frac{\lambda^\mu}{\Lambda_n}\right)\beta_{im}^{(1)}(t)\,\frac{\bar{f}_1(t)}{R_1(t)^{\tau_{i1}}}d\lambda dt. \tag{113}$$

Inasmuch as $r_1(t)$ is positive everywhere on $(0, 1)$ the integrand of (113) is integrable with respect to t by virtue of hypothesis (110). Hence by Lemma I all the integrals (113) for which $m < \tau_{i1}$ are of the order $o(1)$. If $\kappa_2 - \kappa_3 - \delta_{i2} < \tau_{i1}$, this accounts for all the integrals due to the first sum in (111). However, if $\kappa_2 - \kappa_3 - \delta_{i2} = \tau_{i1}$, the integral (113) for which $m = \tau_{i1}$ remains to be considered. A simple extension of the Lemmas III and V to account for the factor $P(\lambda^\mu/\Lambda_n^\mu)$ in the integrand suffices for this consideration, and we find that the integral converges to a value

$$C_{i1}^{(0)} \lim_{t \to 0}\left\{\frac{\bar{f}_1(t)}{t^{\tau_{i1}}}\right\},$$

if $\sigma_i > \tau_{i1}$, or in the case $\sigma_i = \tau_{i1}$, provided the function $\bar{f}_1(t)/t^{\tau_{i1}}$ is of bounded variation in some neighborhood of $t = 0$. In this result $C_{i1}^{(0)}$ is a constant determined by the differential system and is independent of the functions $f_i(t)$.

A similar discussion leads to the evaluation of the integrals arising from the terms of the second sum in (111) and we obtain in this way the result that under the hypotheses made

$$\mathcal{T}_n{}^0(1)\cdot \to \left(C_{i1}^{(0)} \lim_{t \to 0}\left\{\frac{\bar{f}_1(t)}{t^{\tau_{i1}}}\right\} + C_{i2}^{(0)} \lim_{t \to 0}\left\{\frac{\bar{f}_2(t)}{t^{\tau_{i2}}}\right\}\right)\cdot = \lim_{t \to 0}\left(\frac{C_{ij}^{(0)}}{t^{\tau_{ij}}}\right)\bar{\mathfrak{F}}(t)\cdot\,.$$

The discussions of the integrals in the expressions $\mathcal{J}_n^l(1)\cdot$ for $l=1, 2, 3$ follow on similar lines and we obtain thus as a total result, since by formula (59), $\mathcal{A}(1)\cdot = \mathfrak{F}(1)\cdot$, the result

$$,\mathcal{J}_n(1\mathfrak{F}\cdot) \to \mathfrak{F}(1)\cdot + \lim_{t\to 0}\left(\frac{C_{ij}}{t^{\tau_{ij}}}\right)\bar{\mathfrak{F}}(t)\cdot .$$

We see, therefore, that the expansion is summable at $x=1$ if the components of the vector $\bar{\mathfrak{F}}(t)\cdot$ vanish to a sufficiently high degree at $t=0$. In quite the same manner we should find that the expansion is summable at $x=0$ if the components of the given vector vanish to a prescribed degree at $x=1$.

A special case of importance is that in which the numbers τ_{ij} are all zero, for in this case the final term of (114) is a constant matrix multiplied by $\mathfrak{F}(0)\cdot$ and this is $\mathcal{O}$ by formula (58). Hence the expansion converges then at $x=1$ to the value $\mathfrak{F}(1)\cdot$. If in accordance with Theorem 1 the expansion actually converges, then since $\sigma_i \geqq \tau_{ij}$ it follows that the constants τ_{ij} are zero and, moreover, that the same is true of the corresponding exponents associated with the discussion of the expansion at $x=0$. Hence the expansion then converges also at both end points to the respective values of $\mathfrak{F}(x)\cdot$.

The discussion in the case $\alpha=\pi/2$ is similar to that given in detail above, and the results differ only in that the expansion is summable at $x=1$ if the components of the vector $\bar{\mathfrak{F}}(t)\cdot$ vanish to a prescribed degree at $t=1$, while summability at $x=0$ is assured by the vanishing of $\bar{\mathfrak{F}}(t)\cdot$ at $t=0$. In the case $0<\alpha<\pi/2$ a combination of the phenomena described occurs and the proof of summability at either end point is found to be dependent upon the suitable vanishing of the vector $\bar{\mathfrak{F}}(t)\cdot$ at both end points of the interval. We summarize these observations roughly as follows:

THEOREM 7. *If the expansion under the conditions of the respective theorem above is convergent (as distinguished from summable) at the interior points of the interval, then it is also convergent to the value of the given vector at the end points. If the expansion is summable but not convergent at the inner points, then it is summable at the end points also provided the vector $\bar{\mathfrak{F}}(t)\cdot$ vanishes at $t=0$ and $t=1$ to certain degrees prescribed by the differential system.*

UNIVERSITY OF WISCONSIN,
MADISON, WIS.

ON THE DEGREE OF CONVERGENCE OF EXPANSIONS IN AN INFINITE INTERVAL*

BY

W. E. MILNE

In a recent investigation† of the degree of convergence of the Gram-Charlier series in the infinite interval $-\infty < x < \infty$ the writer has shown that in general the convergence is very much less rapid than in the case of a Fourier series under similar circumstances. The question arises whether the slow rate of convergence of this particular series is a characteristic of all such expansions on an infinite interval, or whether there may exist series of orthogonal functions for which the rate of convergence is as rapid as that of the Fourier series in a finite interval. A study of this question not only is of some theoretical interest, but may conceivably serve a very practical end by leading to the discovery of series better adapted to the representation of frequency functions than is the slowly convergent Gram-Charlier series.

This paper is devoted to those expansions on the infinite interval which are associated with the differential equation

$$d^2u/dx^2 + [\lambda - q(x)]u = 0, \tag{1}$$

in which $q(x)$ is real and continuous for all values of x, and

$$\lim_{x=\pm\infty} q(x) = +\infty. \tag{2}$$

The differential equation above is a special case of equations investigated on the infinite interval by Weyl,‡ Hilb,§ Gray,‖ and Milne.¶ It has been shown that there exists an infinite set of critical values of λ, λ_0, λ_1, λ_2, $\cdots$ with limit point at $+\infty$ only, corresponding to which equation (1) has solutions $U_0(x)$, $U_1(x)$, $U_2(x)$, $\cdots$, satisfying the conditions

$$\lim_{x=\pm\infty} U_n(x) = \lim_{x=\pm\infty} U_n'(x) = 0. \tag{3}$$

---

* Presented to the Society, August 27, 1929; received by the editors May 17, 1929.

† These Transactions, vol. 31 (1929), pp. 422–443.

‡ Weyl, Mathematische Annalen, vol. 68 (1910), p. 220, and Göttinger Nachrichten, 1910, p. 442.

§ Hilb, Mathematische Annalen, vol. 76 (1915), p. 333.

‖ Gray, American Journal of Mathematics, vol. 50 (1928), p. 431.

¶ Milne, these Transactions, vol. 30 (1929), p. 797.

The solution $U_n(x)$ vanishes exactly n times, and the integrals

$$\int_{-\infty}^{\infty} U_n^2(x)dx, \quad \int_{-\infty}^{\infty} U_n'^2(x)dx, \quad \int_{-\infty}^{\infty} q(x)U_n^2(x)dx$$

all exist. Aside from the solutions $U_n(x)$ (and constant multiples of them) the equation (1) possesses no solution whose square is integrable from $-\infty$ to $+\infty$.

The functions $U_n(x)$ have the orthogonal properties

$$\int_{-\infty}^{\infty} U_m(x)U_n(x)dx = 0, \qquad m \neq n,$$

so that for an arbitrary function $f(x)$, such that

$$\int_{-\infty}^{\infty} f^2(x)dx$$

exists, we have the formal expansion

$$f(x) = A_0U_0(x) + A_1U_1(x) + A_2U_2(x) + \cdots, \tag{4}$$

in which

$$A_n = \int_{-\infty}^{\infty} f(x)U_n(x)dx \Big/ \int_{-\infty}^{\infty} U_n^2(x)dx. \tag{5}$$

The Gram-Charlier series is a particular case of such an expansion where the function $q(x)$ is a polynomial of the second degree in x having the coefficient of x^2 positive.

It will be shown that the degree of convergence of the series (4) depends upon the function $q(x)$, and that by a suitable choice of $q(x)$ a degree of convergence may be obtained which is only slightly less than that of the Fourier series. It is further shown that the *same* degree of convergence as in the case of the Fourier series is not to be expected.

1. Before we can deal with the series (4) it is necessary to assemble a number of facts regarding the solutions of equation (1). For this purpose we impose some additional restrictions on $q(x)$, and make the following assumptions:

(a) The function $q(x)$ has continuous derivatives of the first three* orders for all values of x.

(b) $$q''(x) \geqq 0, \quad -\infty < x < \infty.$$

* The existence of a continuous third derivative is not necessary, but is convenient for purposes of proof.

There is then no loss of generality in assuming that the origin is so chosen that

$$q(0) = q'(0) = 0.$$

We assume further that

(c)
$$q'(x) > 0, \quad q'''(x) \geqq 0, \quad \text{if} \quad x > 0,$$
$$q'(x) < 0, \quad q'''(x) \leqq 0, \quad \text{if} \quad x < 0;$$

(d)
$$\lim_{x=\pm\infty} q''(x)[q'(x)]^{-4/3} = 0.$$

Now let $u_1(x)$ and $u_2(x)$ denote two particular solutions of (1) satisfying the conditions

(6)
$$u_1(0) = \lambda^{-1/4}, \qquad u_2(0) = 0,$$
$$u_1'(0) = 0, \qquad u_2'(0) = \lambda^{1/4},$$

and form the one-parameter family of solutions

(7)
$$u(x,\theta) = u_1(x)\cos\theta + u_2(x)\sin\theta.$$

The equation of the envelope of the family (7) is

(8)
$$y_{en} = \pm (u_1^2 + u_2^2)^{1/2}.$$

The equation of the locus of the extrema of $u(x, \theta)$ is

(9)
$$y_{ex} = \pm (u_1'^2 + u_2'^2)^{-1/2}.$$

Likewise the equations of the envelope and locus of extrema of $u'(x, \theta)$ are respectively

(10)
$$y_{en}^{(1)} = \pm (u_1'^2 + u_2'^2)^{1/2},$$

and

(11)
$$y_{ex}^{(1)} = \pm (u_1^2 + u_2^2)^{-1/2}.$$

Let us for sake of brevity write

$$u = u(x,\theta), \qquad g^2 = \lambda - q(x),$$

in the interval in which

$$\lambda > q(x),$$

and form the two quadratic expressions*

$$Q_1 = gu^2 + g^{-2}g'uu' + g^{-1}[1 - E]u'^2,$$
$$Q_2 = g^{-1}u'^2 + g^{-2}g'uu' + g[1 + E]u^2,$$

* The use of such forms was suggested by Watson in connection with Bessel functions. Watson, Proceedings of the London Mathematical Society, (2), vol. 16 (1917), pp. 170–171.

in which

$$E = -g^{-3}g''/2 + g^{-4}g'^2.$$

By differentiation and substitution from (1) we have

$$Q_1' = -u'^2(d/dx)(g^{-1}E), \qquad Q_2' = u^2(d/dx)(gE).$$

When the values of these derivatives are calculated in terms of $q(x)$ and reference is made to assumptions (b) and (c), it turns out that Q_1' is negative and Q_2' is positive in the interval for which $x>0$, $\lambda>q(x)$. Therefore if ξ_i denotes a positive root of u' in this interval (and consequently an extremum of u) we have the inequality

$$gu^2(\xi_i) < Q_1(0).$$

Using (6) and (7), together with the equations $q(0)=q'(0)=0$, we find that

$$\begin{aligned} Q_1(0) &= 1 - q''(0)\sin^2\theta/(4\lambda^2), \\ &= 1 + O(\lambda^{-2}). \end{aligned}$$

We shall adopt the notation 1_k to designate any expression of the form $1-O(\lambda^{-k})$. Using this notation and referring to (9) and the definition of $g(x)$, we arrive at the useful result

$$u_1'^2 + u_2'^2 > 1_2[\lambda - q(x)]^{1/2}. \tag{12}$$

Continuing this line of reasoning, we set $u=0$ in Q_1 and then $u'=0$ in Q_2 and again $u=0$ in Q_2 and finally by the aid of (9) and (11) establish three more inequalities

$$u_1'^2 + u_2'^2 < 1_2[\lambda - q(x)]^{1/2}[1 + E], \tag{13}$$

$$u_1^2 + u_2^2 < 1_2[\lambda - q(x)]^{-1/2}, \tag{14}$$

$$u_1^2 + u_2^2 > 1_2[\lambda - q(x)]^{-1/2}[1 - E]. \tag{15}$$

From (13) and (14) we see that

$$\begin{aligned} |u(x,\theta)| &\leqq 1_2[\lambda - q(x)]^{-1/4}, \\ |u'(x,\theta)| &\leqq 1_2[\lambda - q(x)]^{1/4}[1 + E]^{1/2}, \end{aligned}$$

not merely at extrema but for all intermediate values. When $\lambda=\lambda_n$ there is a value of θ, $\theta=\theta_n$, for which

$$u(x,\theta_n) \equiv U_n(x),$$

so that we have at once

$$|U_n(x)| \leqq 1_2[\lambda_n - q(x)]^{-1/4}, \tag{16}$$

$$|U_n'(x)| < 1_2[\lambda_n - q(x)]^{1/4}[1 + E]^{1/2}. \tag{17}$$

The formulas (12)–(17) have been derived for positive values of x. They evidently hold for negative values also as long as $\lambda > q(x)$.

2. Our next object is the determination of a formula connecting λ_n with n. This can be accomplished by means of the fact that the function $U_n(x)$ corresponding to λ_n has exactly n roots $\rho_1, \rho_2, \cdots, \rho_n$, all in the interval for which $\lambda_n > q(x)$. We already know that $U_n(x)$ has exactly n roots, and we now show that $y = U_n(x)U_n'(x)$ does not vanish if $\lambda_n < q(x)$. For by differentiation and use of (1) we get

$$y' = U_n'^2(x) + [q(x) - \lambda_n]U_n^2(x),$$

so that y' is positive if $q(x) > \lambda_n$. Therefore, because of (3), y cannot vanish for any finite x such that $q(x) > \lambda_n$.

The function $U_n(x)$ may be expressed in the form

$$U_n(x) = (u_1^2 + u_2^2)^{1/2} \cos [\phi(x) - \theta_n],$$

where

$$\phi(x) = \arctan (u_2/u_1)$$

and

$$\phi'(x) = (u_1^2 + u_2^2)^{-1}.$$

As x increases from ρ_1 to ρ_n the function ϕ increases by the amount $(n-1)\pi$, so that

$$(n - 1)\pi = \int_{\rho_1}^{\rho_n} (u_1^2 + u_2^2)^{-1} dx.$$

This is the desired formula. By means of (14) and (15) we derive a pair of useful inequalities

$$(18) \qquad (n - 1)\pi > \int_{\rho_1}^{\rho_n} 1_2[\lambda_n - q(x)]^{1/2} dx,$$

$$(19) \qquad (n - 1)\pi < \int_{\rho_1}^{\rho_n} 1_1[\lambda_n - q(x)]^{1/2}[1 - E]^{-1} dx.$$

It is assumed here, and will be proved later, that $E < 1$ in the interval of integration provided λ_n is large.

3. In order to make profitable use of the inequalities just derived it is necessary to learn something about the location of the roots ρ_1 and ρ_n, the location of the largest and smallest roots, ξ_n and ξ_0, of $U_n'(x)$, and to determine the order of magnitude of E. Let h_1 be the negative and h_2 the positive root of the equation

$$q(x) = \lambda_n.$$

Then we know that

$$\rho_n < \xi_n < h_2,$$

and we wish to determine the magnitude of $h_2-\rho_n$.

If we use the notation

$$q_2' = q'(h_2), \qquad q_2'' = (1/2)q''(h_2), \qquad J_2 = q_2' \cdot [q_2']^{-4/3},$$

we see that the inequalities

$$\lambda_n - q(x) \leqq q_2'(h_2 - x),$$
$$\lambda_n - q(x) \geqq q_2'(h_2 - x) - q_2''(h_2 - x)^2$$

are valid for $x<h_2$ because of the assumptions made regarding $q(x)$. The change of variable

$$h_2 - x = [q_2']^{-1/3}t$$

transforms (1) into

$$d^2U_n/dt^2 + \gamma(t)tU_n = 0, \tag{20}$$

where, because of the inequalities above,

$$1 - J_2t \leqq \gamma(t) \leqq 1.$$

By assumption (d) it is possible to make J_2 as small as we please by taking λ_n sufficiently large.

Let t_0 be the value of t corresponding to $x=\rho_n$. Let t_1 be the first positive root of the solution of

$$d^2y/dt^2 + ty = 0 \tag{21}$$

for which $y'=0$, $y\neq 0$ at $t=0$. We find by comparing (21) and (20),

$$t_0 > t_1.$$

Also if t_2 is the first positive root of that solution of (21) for which $y'\neq 0$, $y=0$ at $t=0$, and if λ_n is so large that $1-J_2t_0>0$, then

$$t_0[1 - J_2t_0]^{1/3} < t_2.$$

If we also take λ_n so large that $J_2<1/(3)t_2$, then the last inequality assures us that

$$t_0 < 2t_2.$$

These results give us the inequalities

$$t_1[q_2']^{-1/3} < h_2 - \rho_n < 2t_2[q_2']^{-1/3}. \tag{22}$$

By numerical calculation it is found that $t_1=1.9\cdots$, $t_2=2.8\cdots$.

We are now in a position to estimate the magnitude of the largest maximum of $|U_n(x)|$ for positive values of x. This occurs at the point $x=\xi_n$ and it is therefore apparent from the fact that ξ_n is between ρ_n and h_2 that

$$\text{Max}\,|\,U_n(x)\,| < (h_2 - \rho_n)\,|\,U_n'(\rho_n)\,|\,.$$

Using (17) and (22) we get

$$\text{Max}\,|\,U_n(x)\,| < M_2[q_2']^{-1/6},$$

where

$$M_2 = 1_2(2t_2)^{5/4}[1 + E(\rho_n)]^{1/2}.$$

It remains to investigate E. We find

$$E = (3/8)(\lambda - q(x))^{-3}q'^2(x) + (1/4)(\lambda - q(x))^{-2}q''(x).$$

First of all it is apparent that for x in any fixed finite interval

$$E = O(\lambda^{-2}).$$

Next we note that, because of the hypotheses regarding $q(x)$ and the inequalities (22),

$$E(\rho_n) < (3/8)t_1^3 + \epsilon,$$

where ϵ approaches zero as λ becomes infinite. Therefore, a fortiori,

$$E(x) < (3/8)t_1^3 + \epsilon, \quad 0 < x < \rho_n.$$

All the results of this paragraph have been obtained for x positive. Entirely similar results may be obtained when x is negative. For example when x is negative

$$\text{Max}\,|\,U_n(x)\,| < M_1[-q_1']^{-1/6}.$$

Therefore if δ denotes the smaller of the two quantities $-q_1'$ and q_2', we have

$$\text{Max}\,|\,U_n(x)\,| < M\delta^{-1/6}, \quad -\infty < x < \infty. \tag{23}$$

4. Because of the inequalities just obtained for E it is clear that the integrals in (18) and (19) are of the same order of magnitude, and therefore either one will serve for the determination of the order of magnitude of λ_n in terms of n. Moreover by using the inequalities $\lambda_n - q(x) < q_2'(h_2 - x)$, and (22) we find that

$$\int_{\rho_n}^{h_2} (\lambda_n - q(x))^{1/2}dx < (2/3)(2t_2)^{3/2},$$

with a similar result for the integral from h_1 to ρ_1, so that for the determination of the order of magnitude of λ_n we may use the approximate equation

$$(n-1)\pi = \int_{h_1}^{h_2} (\lambda_n - q(x))^{1/2}dx \qquad \text{(approx.)}. \tag{24}$$

It is interesting to test this equation for a problem in which the value of λ_n is known beforehand, namely where $q(x) = -1/2 + x^2/4$. Here we know that $\lambda_n = n$. For this case the value of the integral in (24) proves to be $\pi(\lambda_n + 1/2)$, giving us $\lambda_n = n - 3/2$, a rather surprisingly good approximation.

The integral in (24) may be transformed as follows: Let $x = h_2(z)$ be the inverse of $z = q(x)$ when x is positive and let $x = -h_1(z)$ be the inverse when x is negative. In the integral from h_1 to 0 we make the substitution $x = -h_1(\lambda_n s)$ and in the integral from 0 to h_2 the substitution $x = h_2(\lambda_n s)$, then integrate each term by parts, and obtain

$$2\pi(n-1) = \lambda_n^{1/2}\left\{\int_0^1 \frac{h_2(\lambda_n s)ds}{(1-s)^{1/2}} + \int_0^1 \frac{h_1(\lambda_n s)ds}{(1-s)^{1/2}}\right\} \tag{25}$$

approximately. Since $h_1(\lambda_n s)$ and $h_2(\lambda_n s)$ become infinite with λ_n, we have from (25)

THEOREM I. *The order of magnitude of λ_n is less than n^2.*

On the other hand if $q(x) = x^{2\kappa}$, we find directly from (25) that λ_n is of the order of magnitude of $n^{2-2/(\kappa+1)}$. By taking κ large enough we may make the exponent as nearly 2 as we please. In fact if $q(x)$ behaves like $e^{|x|}$ for x large, say $q(x) = \cosh x - 1$, we find that the order of magnitude of λ_n is at least as great as $n^2/\log^2 n$. Further refinements may be made indefinitely. Hence we have proved

THEOREM II. *By a suitable choice of $q(x)$ the order of magnitude of λ_n may be made as near n^2 as we please.*

5. It is necessary to obtain a lower bound for the integral in the denominator of (5). Now as we go farther from the origin in either direction the areas under the arches of the curve $y = U_n^2(x)$ increase, and there are in all $(n+1)$ arches. Hence the value of the integral will be greater than $(n+1)a_0$, where a_0 is the area of the smallest arch. Since the smallest maximum of $U_n^2(x)$ is greater than or equal to $1_2 \cdot \lambda_n^{-1/2}$, and the smallest distance between a maximum and a root is greater than $\pi/(2\lambda_n^{1/2})$, we find that $a_0 > \pi/(3\lambda_n)$. Therefore

$$\int_{-\infty}^{\infty} U_n^2(x)dx > \pi(n+1)/(3\lambda_n). \tag{26}$$

6. We are now ready to take up the question of convergence of (4) and enunicate the first result as follows:

THEOREM III. *If $f(x)$ has a continuous derivative of bounded variation in the infinite intervals; if the integral*

$$\int_{-\infty}^{x} f(s)q(s)ds$$

exists and is bounded; and if the Stieltjes integral

$$I = \int_{-\infty}^{\infty} [1 + q(x)]^{1/4} |dF(x)|$$

exists, where

$$F(x) = f'(x) - \int_{-\infty}^{x} f(s)q(s)ds\ ;$$

then

$$R_n(x) = \sum_{i=n+1}^{\infty} A_i U_i(x) = O(\delta^{-1/6}\lambda_n^{-1/4}),\quad -\infty < x < \infty\ ,$$

$$= O(\lambda_n^{-1/2}),\quad \alpha < x < \beta.$$

Multiply the equation

$$U_n(x) = \lambda_n^{-1} q(x) U_n(x) - \lambda_n^{-1} U_n''(x)\ , \tag{27}$$

which is in effect equation (1), by $f(x)dx$, integrate from $-\infty$ to $+\infty$, and integrate each term on the right by parts. The result is

$$\int_{-\infty}^{\infty} f(x)U_n(x)dx = \lambda_n^{-1} \int_{-\infty}^{\infty} F(x)U_n'(x)dx,$$

since the integrated terms vanish at the limits because of (3). Now $U_n(x)$ is continuous and $F(x)$ is of bounded variation, so that we may integrate by parts between finite limits α and β, and obtain

$$\int_{\alpha}^{\beta} F(x)U_n'(x)dx = F(\beta)U_n(\beta) - F(\alpha)U_n(\alpha) - \int_{\alpha}^{\beta} U_n(x)dF(x)\,.$$

Because of the hypothesis regarding the existence of the Stieltjes integral on the infinite interval we may let α and β become infinite, and obtain finally

$$\int_{-\infty}^{\infty} f(x)U_n(x)dx = -\lambda_n^{-1} \int_{-\infty}^{\infty} U_n(x)dF(x)\,.$$

We easily see that the right hand integral is less in absolute value than

$$\lambda_n^{-1}\ \mathrm{Max}\,|\,U_n(x)\,|\,[1 + q(x)]^{-1/4} I,$$

and because of (16) and (23) the whole expression is $O(\lambda_n^{-5/4})$. Referring now to (5) and (26) we have

$$A_n = O(\lambda_n^{-1/4} n^{-1}).$$

Therefore

$$\begin{aligned} A_n U_n(x) &= O(\delta^{-1/6} \lambda_n^{-1/4} n^{-1}), \quad -\infty < x < \infty, \\ &= O(\lambda_n^{-1/2} n^{-1}), \quad \alpha < x < \beta, \end{aligned}$$

and

$$\begin{aligned} R_n(x) &= O(\delta^{-1/6} \lambda_n^{-1/4}), \quad -\infty < x < \infty, \\ &= O(\lambda_n^{-1/2}), \quad \alpha < x < \beta. \end{aligned}$$

For the case in which $q(x) = x^{2\kappa}$ we have

$$\delta = O(\lambda_n^{1-1/\kappa})$$

and

$$\lambda_n = O(n^{2-2/(\kappa+1)}),$$

so that the above equations become

$$\begin{aligned} R_n(x) &= O(n^{(-5\kappa+1)/(6\kappa+6)}), \quad -\infty < x < \infty, \\ &= O(n^{-\kappa/(\kappa+1)}), \quad \alpha < x < \beta. \end{aligned}$$

For $\kappa = 1$ we have the same results that were secured for the Gram-Charlier series in the paper referred to. For large values of κ the exponents are approximately $-5/6$ and -1 respectively. In view of Theorem I it seems that we need hardly expect actually to attain these values for the class of expansions under discussion.

THEOREM IV. *If $f(x)$ has a continuous second derivative of bounded variation on the infinite interval; if the Stieltjes integral*

$$I_1 = \int_{-\infty}^{\infty} [1 + q(x)]^{3/4} \, | \, dF_1(x) \, |$$

exists, where

$$F_1(x) = q(x) f(x) - f''(x);$$

and if

$$\lim_{x = \pm\infty} F_1(x) = 0;$$

then

$$\begin{aligned} R_n(x) &= O(\delta^{-1/6} \lambda_n^{-3/4}), \quad -\infty < x < \infty, \\ &= O(\lambda_n^{-1}), \quad \alpha < x < \beta. \end{aligned}$$

Multiply the equation (27) by $f(x)dx$, and integrate from $-\infty$ to $+\infty$, integrating the second term on the right by parts twice. Then

$$\int_{-\infty}^{\infty} f(x) U_n(x) dx = \lambda_n^{-1} \int_{-\infty}^{\infty} F_1(x) U_n(x) dx.$$

As in the proof of Theorem III we again integrate by parts to obtain

$$\int_{-\infty}^{\infty} f(x)U_n(x)dx = -\lambda_n^{-1}\int_{-\infty}^{\infty}\left[\int_0^x U_n(s)ds\right]dF_1(x).$$

By exactly the same process as was used in the paper on Gram-Charlier series* it may be shown that

$$\int_0^x U_n(x)dx = O(\lambda^{-5/4})$$

in any fixed finite interval and is bounded in all intervals. With this fact at hand we may complete the proof in substantially the same manner as in Theorem III.

For the case of a function $f(x)$ with a continuous kth derivative of bounded variation in the infinite interval, the generalization of Theorems III and I is easily accomplished. For we have

$$\int_{-\infty}^{\infty} f(x)U_n(x)dx = \lambda_n^{-m}\int_{-\infty}^{\infty} F_m(x)U_n(x)dx,$$

where

$$F_1(x) = q(x)f(x) - f''(x),$$
$$F_2(x) = q(x)F_1(x) - F_1''(x),$$
$$\cdots\cdots\cdots\cdots$$

and where $m=k/2$ if k is even, and $m=(k-1)/2$ if k is odd. For the case where k is odd the proof is completed as in Theorem III, while if k is even it is completed as in Theorem IV. The formulation and proof of this generalized theorem may be left to the reader.

The foregoing arguments establish the uniform convergence of the series (4) in the infinite interval but do not of themselves prove that $f(x)$ is equal to the sum of the series. This fact has been established by Weyl† on the assumption that

$$\int_{-\infty}^{\infty}[q(x)f(x) - f''(x)]^2dx$$

exists.

In cases where Weyl's result is not directly applicable, as for example in the case of Theorem III, it is still easy to draw the desired conclusion. For either by following Weyl's reasoning or by applying Hilbert's theory of

* Loc. cit., paragraph 4, pp. 427–429.

† Göttinger Nachrichten, 1910, pp. 449–450.

integral equations* directly to our particular problem we may show that the set of functions $U_0(x)$, $U_1(x)$, $U_2(x)$, $\cdots$ is *closed*. The convergence to $f(x)$ is therefore a consequence of uniform convergence.

In conclusion it should be pointed out that the results of this paper may be extended to cases where $q(x)$ does not fulfil the conditions (b) and (c) in the entire infinite interval, provided these conditions hold for $|x|$ sufficiently large, say $|x|>X$. For when λ_n is sufficiently large we have an asymptotic representation of $U_n(x)$ of the form

$$U_n(x) = \lambda_n^{-1/4} \sin \lambda^{1/2}(x - x_0) + O(\lambda^{-3/4}),$$

valid in the interval $-X<x<X$, from which upper and lower bounds for the number of zeros in this interval may be obtained, together with upper bounds for $|U_n(x)|$, $|U_n'(x)|$ and $\int_0^x U_n(x)dx$. Outside of this interval a suitable modification of the treatment given here may be applied.

Having thus determined the orders of magnitude of λ_n, $|U_n(x)|$, $|U_n'(x)|$ and $\int_0^x U_n(x)dx$, we repeat the proofs of Theorems III and IV without change.

---

* Hilbert, Göttinger Nachrichten, 1904, pp. 73–74 and pp. 221–222.

STANFORD UNIVERSITY,
STANFORD UNIVERSITY, CALIF.

A KINEMATICAL TREATMENT OF SOME THEOREMS ON NORMAL RECTILINEAR CONGRUENCES*

BY

CHARLES H. ROWE

1. **Introduction.** When we are discussing the motion of a rigid body in space, we often have occasion to consider the ruled surfaces which are described in space and in the body by the same moving straight line, or, what amounts to the same thing, the surface which a moving straight line actually describes during its motion, and the surface which it appears to describe to an observer who is moving at the same time in an assigned manner. Usually the motions with which we are concerned are ordinary or one-parameter motions, but an analogous idea presents itself at once when we consider instead two-parameter motions. Thus, with each of the ∞^2 positions of a straight line l which has a two-parameter motion in space, we may associate a position of a rectangular trihedral T, and we may imagine an observer (whom also we shall call T) who shares the two-parameter motion of this trihedral and uses it as his system of reference, estimating each position of the straight line l by referring it to the corresponding trihedral T. We can then consider two rectilinear congruences: that which the line l actually describes in space during its two-parameter motion, and that which it appears to describe to the observer T.† The second congruence is of course equivalent to that which is described by a line l' whose position relative to a trihedral fixed in space is the same as the position of l relative to the corresponding trihedral T.

Instead of a straight line which describes a congruence, we may, more generally, consider a geometrical element e of any nature, and if, as before, we associate a trihedral T with each of the ∞^2 elements e and personify it by an observer T, we can speak of the two-dimensional manifold of elements which e really describes and that which it appears to describe to the observer T. It is clearly not necessary, however, that each manifold should contain ∞^2 distinct elements; it may happen, for instance, that, while the element e assumes ∞^2 distinct positions in space, the observer T may regard it as assuming only ∞^1 distinct positions or even as being fixed.

---

* Presented to the Society, August 30, 1929; received by the editors in July, 1928.

† A pair of congruences which are related in this manner are considered by Darboux. (*Leçons sur la Théorie Générale des Surfaces*, vol. 4, p. 125.)

I propose to prove in this paper that, under certain conditions, a straight line which appears to the observer T to describe a normal congruence does so in reality. I shall then show that certain known theorems on normal congruences can be derived from this result in a simple manner without analysis and that in some cases extensions which I believe to be new suggest themselves.

2. **Kinematical preliminaries.** When infinitesimal displacements are given to a trihedral T and to a point P, there are three displacements associated with the point P:

(i) the actual displacement of the point P;

(ii) the displacement which the observer T attributes to the point P, or, more briefly, the apparent displacement of P;

(iii) the displacement of a point which is fixed with reference to T and which coincides with P when T and P are in their initial positions, or, as we shall call it, *the displacement of the moving system at* P.

We shall recall the familiar fact that the actual displacement of P is the resultant of the apparent displacement of P and the displacement of the moving system at P.

Let us now suppose that a two-parameter motion of a trihedral T has been defined by associating a position of T with each pair of values of two parameters u and v. Corresponding to each position (u, v) of the trihedral there are certain straight lines which we shall call *special lines*, and which we may define by saying that a line l is a special line if, in every infinitesimal displacement of T from the position (u, v) to an adjacent position $(u+du, v+dv)$, the displacement of the moving system at each point on l is normal to l.

In order to determine these lines we shall make use of the fact that, corresponding to any position (u, v) of T, there exists in general a pair of straight lines, which we shall call the *axes*, such that T may be brought from its initial position (u, v) to any adjacent position $(u+du, v+dv)$ by a combination of rotations of suitable magnitudes about these two axes.* It is then easy to see that the special lines are the ∞^2 common transversals of the two axes. The axes may of course be imaginary or coincident, but there are always ∞^2 real special lines forming a linear congruence.

There is one special case which we shall have to consider: that in which the axes are indeterminate. When this happens every infinitesimal displacement of the trihedral reduces to a rotation about some straight line passing through a certain fixed point and lying in a certain fixed plane through this

* See Mannheim, *Principes et Développements de Géométrie Cinématique*, 1894, p.127, or Darboux, *Leçons*, vol. 1, p. 91.

point. The special lines then consist of all lines passing through the point together with all lines lying in the plane. The point and the plane may be both at infinity, and in this case every infinitesimal displacement is a translation parallel to a certain plane, and the special lines that are not wholly at infinity are the lines which are perpendicular to this plane.

Another exceptional case is that in which the trihedral takes only ∞^1 distinct positions in spite of the fact that we have used two parameters to specify each of its positions. If this happens, there are ∞^3 special lines corresponding to each position of the trihedral, and they form a linear complex. We shall assume in what follows that this case does not arise.

3. **Normal congruences.** Considering again the two-parameter motion of a straight line l and the associated trihedral T, we can state our fundamental result as follows:

If in each of its positions the line l is a special line with reference to the corresponding position of the trihedral T, then l will describe a normal congruence in space during its two-parameter motion if, and only if, the observer T regards it as describing a normal congruence. Further, when l does describe a normal congruence, a point on l which appears to describe a surface cutting all the rays l at right angles does so in reality.

We shall suppose that the straight line l assumes ∞^2 distinct positions both in reality and as viewed by T, and that the congruence which it appears to describe is a normal congruence. We can then find on each ray l a point P such that every apparent infinitesimal displacement of P is normal to l. If we take any position of the ray l and consider the corresponding positions of the trihedral T and the point P, it is clear that, in any infinitesimal displacement from this position, the apparent displacement of P and the displacement of the moving system at P are both normal to l. Hence the actual displacement of P, which is their resultant, is also normal to l, or, in other words, the locus of the point P in space is a surface which cuts all the rays at right angles.

We may remark that, if we have constructed a normal surface S of the congruence which l actually describes, and if T observes this surface from each of his positions, he will believe that it moves with a two-parameter motion and envelops a surface S', and S' will be for T a normal surface of the congruence which he believes the ray l to describe. If T were to construct this surface S' and to fix it in position in his trihedral, then, from the point of view of a fixed observer, S' would move and would envelop S.

We have omitted from our proof the case in which one or other of the two congruences under consideration reduces to a system of only ∞^1 distinct lines or to a single line. If this happens, the other congruence will be a normal

congruence without further conditions beyond the requirement that each ray should be a special line. We have only to observe that, in the case of a ruled surface or of a single straight line, we can always find a point P on each line such that every displacement of P is normal to the line, remembering that we are entitled to regard a displacement of zero magnitude as being normal to any direction that we please. We shall therefore agree to say that a line describes a normal congruence when it assumes at most ∞^1 distinct positions.

Our theorem admits the following converse, the proof of which will be evident:

If a ray l describes a normal congruence both in reality and in appearance, and if we can find a point P on the ray in each of its positions so that P both in reality and in appearance describes a surface which cuts all the rays at right angles, then each ray is a special line with reference to the corresponding trihedral.

Our theorem and its converse may be thrown into an alternative form. Let us associate with each position of the moving trihedral a plane element formed by a point of a certain fixed surface and the tangent plane thereat. This element will appear to T to describe a manifold of ∞^2 plane elements which in general is not formed by the elements of any surface. *In order that this manifold should consist of the elements of some surface, it is necessary and sufficient that the line drawn through the point of each plane element perpendicular to its plane should be a special line with reference to the corresponding position of T.*

4. **The theorems of Beltrami and Ribaucour.** We shall now point out that, when we take into account a certain elementary property of the motion of a surface which rolls upon an applicable surface, the classical theorems of Beltrami and of Ribaucour on the deformation of congruences may be regarded as almost immediate consequences of our result.*

Consider a pair of applicable surfaces S and S' which are not both ruled surfaces, and let S' be fixed in space while S rolls upon it so that each point M of S is brought in turn into coincidence with the corresponding point M' of S', the tangent planes, and corresponding directions in the tangent planes, also coinciding. Let T be any trihedral invariably connected with the rolling surface S, and, with each position of T, let us associate (a) a ray l passing through the corresponding point of contact of the two surfaces, (b) a ray m lying in the common tangent plane at the point of contact. Now,

* Proofs of these theorems depending on kinematical considerations are given by Darboux (*Leçons*, vol. 3, pp. 348, 352), but they are of a different nature from ours.

if we consider the congruence which the ray l really describes and subject it to a deformation in Beltrami's sense corresponding to the deformation of S' into S, we shall clearly obtain a congruence which is identical with that which the ray l appears to describe to the observer T. This remark, together with an analogous remark about the ray m, shows that, in order to prove the theorems of Beltrami and Ribaucour, we need only assure ourselves that the rays l and m are both special lines with reference to the corresponding position of T. This, however, is evident when we remember that every infinitesimal displacement of the rolling surface is a rotation about some line touching both surfaces at their point of contact, so that, in virtue of a remark which we have already made, the special lines consist of all lines passing through the point of contact, together with all lines lying in the common tangent plane at the point of contact.

It will be clear that the proposition given at the close of the preceding section leads in a similar manner to a theorem due to Bianchi* which may be stated as follows: Let us associate a plane element e with each point M of a surface S. Let S be deformed and, as it varies, let each of the plane elements move so that it retains an invariable position with reference to the point M and to elements of the surface issuing from M. If we suppose that these plane elements were those of some surface Σ in their original positions, then, in order that they should still be the plane elements of some surface after deformation, it is necessary and sufficient that each normal to Σ should either pass through the corresponding point M of S, or else lie in the tangent plane at M to S.

We may also remark that Ribaucour's theorem on the deformation of a normal congruence of plane curves lying in the tangent planes of a surface can be derived in like manner from our result.

5. **Another theorem of Ribaucour.** Our next application will yield a result which, as we shall see, may be regarded as a generalisation of another theorem of Ribaucour.

Consider any two surfaces S, S' between whose points M, M' a correspondence of any nature has been established, and construct the surface Σ which is the locus of the extremity P of a vector OP drawn through a fixed origin O equal and parallel to MM'. With the points M and M' of the two surfaces associate rays m and m' respectively, both parallel to the normal to Σ at the corresponding point P, and so situated that by subjecting them to a common translation we could cause m to pass through M and m' to pass through M'. We shall show that, as M and M' describe the surfaces on which

* Rendiconti dei Lincei, (5), vol 24 (1915), p. 4.

they lie, *neither of the two rays can describe a normal congruence unless the other does so also.*

We shall hold S' fixed in space, and we shall attach a trihedral to S and allow S to move with a two-parameter motion, carrying its associated system of rays invariably fixed to it, so that corresponding points of the two surfaces are in turn brought into coincidence, and so that every position of S may be derived from its original position by a translation. It is easy to see that all infinitesimal displacements from the position in which M and M' coincide are translations parallel to the tangent plane to Σ at the corresponding point P, so that any line parallel to the normal to Σ at P is a special line. Now, when M and M' coincide, the corresponding rays m and m' coincide in a single line which is a special line for the corresponding position of the trihedral. If we identify this line with the line l of our fundamental result, the truth of our assertion will be evident.

We shall now particularise the nature of the correspondence between the two surfaces and we shall suppose that it is a correspondence by parallel tangent planes, so that we can find corresponding nets of curves on the two surfaces such that corresponding curves have the same direction at corresponding points (and in general this can be done in one and only one way). If, further, we note that the tangent planes at the points M, M', P to the respective surfaces are now all parallel, it will be clear that our result may be stated as follows:

If we have traced corresponding nets on two surfaces S and S' so that corresponding tangents to curves of the nets are parallel, and if with corresponding points M, M' of the two surfaces we associate rays m, m' parallel to the common direction of the normals at M and M' in such a way that the position of m in relation to the tangents to the net at M is the same as the position of m' in relation to the tangents to the net at M', then, if one of these rays describes a normal congruence, the other does so also.

The theorem of Ribaucour* to which we have referred need not be enunciated separately, for it is the particular case of what we have just stated which arises when the surfaces have the same spherical representation of their lines of curvature, the nets which we have introduced being identified with the lines of curvature.

6. **The moving trihedral of Darboux.** The remainder of our work arises out of two theorems which were given by Caronnet, and, in order to establish and to extend these theorems, it will be necessary to make some remarks on the circumstances which arise when we identify the trihedral T of our initial

* Journal de Mathématiques, (4), vol. 7 (1891), p. 40.

result with the moving trihedral of Darboux. We shall thus associate with each point M of a given surface a trihedral formed by three mutually perpendicular straight lines intersecting in M, one of which coincides with the normal to the surface at M. The position of the other two edges of the trihedral in the tangent plane may conveniently be specified by tracing on the surface a singly infinite family of curves Γ and stipulating that the tangent at a point M to the curve Γ that passes through it has an assigned fixed position in relation to the moving trihedral. Since for our purposes two trihedrals are equivalent if one is fixed in relation to the other, we shall not need to know how this assigned fixed position is situated in relation to the trihedral, the fact that it is fixed being all that is relevant. We see, then, that in order to complete the determination of the motion of the trihedral, it is a matter of indifference whether we use a family of curves Γ or a family of curves which cut the curves Γ at an arbitrary constant angle.

The axes a_1 and a_2 of a Darboux trihedral are always real: a_1 is a line in the normal plane of the line of curvature C_1 at M passing through the corresponding center of principal curvature P_1, and a_2 is a line in the normal plane of the line of curvature C_2 at M passing through the corresponding center of principal curvature P_2.* We can easily complete the statement of the position of these axes when the indeterminateness in the motion of the trihedral has been removed by tracing a family of curves Γ; and to make this clear we shall first show that the center of geodesic curvature of the curve Γ which passes through M lies on the straight line which joins the points A_1 and A_2 where the axes a_1 and a_2 cut the tangent plane.

Draw through M the straight line in the tangent plane which is normal to the curve Γ and let it cut the straight line A_1A_2 in G. Consider that infinitesimal displacement of the moving trihedral T in which M moves along Γ. The displacement of the moving system at G is parallel to the normal to the surface at M, for it is perpendicular to A_1A_2, which is a special line, and it is also perpendicular to MG because at M, and hence at every point on MG, the displacement of the moving system is perpendicular to MG. Combining this with the displacement along MG which the observer T attributes to the point G, we see that the actual displacement of G lies in the normal plane of the curve Γ at M.

This fact suffices to show that G is the center of geodesic curvature of the curve Γ,† for it implies that G lies on the characteristic of the normal plane of Γ corresponding to a displacement of M along Γ, and this characteristic

* Darboux, *Leçons*, vol. 2, p. 365.

† Cf. Darboux, *Leçons*, vol. 3, p. 117.

is the polar line of the curve, so that the point G where it meets the tangent plane is the center of geodesic curvature of the curve.

If we trace any system of isogonal trajectories Γ' of the curves Γ, it follows similarly that the center of geodesic curvature of the curve Γ' which passes through M also lies in the line A_1A_2. We may thus construct the line A_1A_2 by joining the centers of geodesic curvature of the curves Γ and Γ' which pass through M. The axes of the trihedral are then found by joining the points where this line cuts the normal planes of the lines of curvature to the corresponding centers of principal curvature. In other words, the axes are the polar lines of those isogonal trajectories of the curves Γ which touch the lines of curvature at M.

Incidentally, we have proved a theorem which is due to Lipka.* We have shown that, if we construct the ∞^2 isogonal trajectories of a singly infinite family of curves on a surface, the locus of the centers of geodesic curvature of the trajectories passing through a point is a straight line.

7. **The theorems of Caronnet.**† The first of these states that the necessary and sufficient condition that the line joining the centers of geodesic curvature of the curves of an orthogonal net on a surface should describe a normal congruence is that the corresponding radii of geodesic curvature should be connected by some functional relation.

This theorem was extended by Beal,‡ who proved that we may replace the orthogonal net by any net the curves of which cut at a constant angle, and showed that the result remains true in the limiting case in which one of the two centers of geodesic curvature is constantly at infinity.

Caronnet's second theorem states that if, corresponding to each point on a surface, we construct the line joining the center of principal curvature of the line of curvature of the first system at the point to the center of geodesic curvature of the line of curvature of the second system at the point, this line will describe a normal congruence if, and only if, the corresponding radii of principal and geodesic curvature satisfy some functional relation.§

To prove Beal's extension of Caronnet's first theorem, we shall suppose that we are given a net of curves formed by a family of curves Γ together with their trajectories Γ' at some constant angle. The line l joining the centers of geodesic curvature at a point M on the surface of the curves Γ and Γ' is then a special line for the trihedral whose motion has been determined by

* These Transactions, vol. 13 (1912), p. 94.

† Comptes Rendus, vol. 115 (1892), p. 589.

‡ American Journal of Mathematics, vol. 35 (1913), p. 24.

§ A proof of this theorem is given by Pell, American Journal of Mathematics, vol. 20 (1898), p. 112.

means of the curves Γ or the curves Γ' in the manner which we have explained. In order that l should describe a normal congruence, it is therefore necessary and sufficient that the observer T should regard it as doing so. Now, for the observer T, l lies in a fixed plane, and he can therefore regard l as describing a normal congruence only if it assumes at most ∞^1 distinct positions in relation to his trihedral. The condition that this should be so is clearly the existence of a functional relation between the radii of geodesic curvature of the curves Γ and Γ', for these radii are the intercepts which the line l makes on two lines which are fixed in one of the faces of the trihedral.

We may notice that, if we are given the functional relation, we can at once derive the normal surfaces of the congruence which l describes from the involutes of the curve which l appears to envelop. For instance, we can see immediately the truth of a theorem given by Beal which asserts that the foot of the perpendicular from M on l describes a normal surface of the congruence if, and only if, the functional relation between the radii of geodesic curvature is that which expresses the constancy of their ratio.

In order to discuss Caronnet's second theorem, we shall identify the curves Γ which we use to define the motion of the trihedral with one of the families of lines of curvature. The axes a_1 and a_2 of the trihedral are now the polar lines of the lines of curvature,* so that a_1 is the line P_1G_1 in the normal plane of the line of curvature C_1 which joins the center of geodesic curvature G_1 of C_1 to the center of principal curvature P_1 corresponding to C_1, and a_2 is the line P_2G_2 which is similarly related to the line of curvature C_2.

The theorem in question follows at once, for P_1G_2 is a special line and, as in the preceding theorem, it appears to lie in a fixed plane. The condition that it should describe a normal congruence is therefore the existence of a functional relation between the radii of principal and geodesic curvature MP_1 and MG_2.

8. **Extensions of Caronnet's second theorem.** Caronnet's second theorem leads to the construction of a normal congruence associated with the lines of curvature of a surface only when the surface is of a special type. Our methods, however, suggest the following propositions which are of a similar nature, but which lead to the construction of a normal congruence whatever be the nature of the surface:

The line G_1Q joining the center of geodesic curvature of C_1 at M to a point Q on the polar line of C_2 describes a normal congruence if Q is so chosen that there exists a functional relation between the length MQ and the radius of geodesic curvature MG_1.

* Darboux, *Leçons*, vol. 3, p. 345.

The line P_1R joining the center of principal curvature P_1 to a point R on the tangent to the corresponding line of curvature C_1 describes a normal congruence if R is so chosen that there exists a functional relation between the length MR and the radius of principal curvature MP_1.

The truth of these propositions will be sufficiently evident if we remark that in each case the straight line under consideration is a special line which appears to the observer T to intersect constantly a fixed straight line, and that the only normal congruences in which every ray intersects a fixed straight line are those whose rays are generators of a system of ∞^1 right circular cones which have this straight line as their common axis.

The second of these two results, however, is equivalent to a theorem of Beltrami,* for, if we look at the matter from the standpoint of the sheet S_1 of the surface of centers on which P_1 lies, and remember that the curves which the normals to the original surface touch on S_1 form a family of geodesics, along whose orthogonal trajectories MP_1 is constant, we see that our result is equivalent to the statement that, if on any surface S_1 we trace a family of geodesic parallels and draw through each point of the surface a ray normal to the parallel and cutting the surface at an angle which is constant along each of the parallels, these rays will form a normal congruence.

The following is another result of a similar character:

If a ray l meets the polar lines of the lines of curvature at R_1 and R_2 and makes a constant angle with the normal to the surface, the condition that it should describe a normal congruence is the existence of a functional relation between the perpendicular distances of R_1 and R_2 from the normal to the surface.

The proof of this proposition will be sufficiently clear if we remark that the only normal congruences in which each ray makes the same angle with a fixed straight line are those whose rays all touch a cylinder having its generators parallel to the fixed straight line.

We may note that this result may be enunciated in a form which involves a reference only to the congruence formed by the normals P_1P_2 and not to a particular normal surface of this congruence. We could, in fact, have equally well described the polar lines of the lines of curvature as the tangent lines to the sheets of the focal surface of the congruence of normals which are conjugate to the ray P_1P_2 of this congruence.† We might also have described them as the rectifying lines of the curves which the rays P_1P_2 touch on the sheets of the focal surface, and, with this alteration in the enunciation, the particular case of our result in which the constant angle

* See Darboux, *Leçons*, vol. 3, p. 349.

† Darboux, *Leçons*, vol. 3, p. 335.

between l and P_1P_2 is a right angle can be extended to congruences which are not normal, but are subject to the condition that the angle between the focal planes is constant. We have in fact the following theorem:

If with each ray m of a congruence in which the angle between the focal planes is constant we associate another ray l which is perpendicular to m and meets the rectifying lines of the curves which the rays m touch on the sheets of the focal surface, the necessary and sufficient condition that the ray l should describe a normal congruence is the existence of a functional relation between the perpendicular distances from m of the points where l cuts the rectifying lines.

Let P_1, P_2 be the foci of the ray m, lying respectively on the sheets S_1, S_2 of the focal surface, and let the rays touch the curves Γ_1 on S_1 and the curves Γ_2 on S_2. With each ray m associate a trihedral T whose motion we shall not specify completely but shall subject only to the condition that the focal planes of the ray m are fixed planes with reference to T. The focal planes are the osculating planes of the curves Γ_1, Γ_2, so that the rectifying planes are the planes through the ray m perpendicular to the focal planes, and are therefore fixed planes with reference to T.

We shall show that the special lines which lie in planes perpendicular to m intersect the two rectifying lines. Take any point M on m and through it draw a plane at right angles to m. Consider the special line which lies in this plane, and let it cut the rectifying planes of the curves Γ_1 and Γ_2 in G_1 and G_2 respectively. If we give to T that infinitesimal displacement in which the focus P_1 moves along Γ_1 and notice that both the actual and apparent displacements of M lie in that focal plane which osculates Γ_1 at P_1, we see that the displacement of the moving system at M also lies in this plane and is therefore normal to MG_1. Arguing as we did in our discussion of the Darboux trihedral, we see that the displacement of the moving system at G_1 is parallel to the ray m, for it is perpendicular to G_1G_2 which is a special line, and it is perpendicular to MG_1 because at M, and hence at every point on MG_1, the displacement of the moving system is perpendicular to MG_1. Combining this displacement of the moving system with the displacement in the rectifying plane which the observer T attributes to the point G_1, we see that the actual displacement of G_1 also lies in this rectifying plane, so that G_1 is a point on the rectifying line of the curve Γ_1. Similarly, G_2 is a point on the rectifying line of the curve Γ_2. Hence those special lines which are perpendicular to the ray m are transversals of the two rectifying lines. It follows then that the line l of our theorem is a special line, and therefore l will describe a normal congruence if, and only if, the observer T regards it as doing so. The truth of our theorem will then be evident if we remember that the condition that a straight line which always remains parallel to a fixed

plane should describe a normal congruence is that its projection on this plane should have only one degree of freedom.

We may remark in conclusion that, in the course of the foregoing argument, we have incidentally shown that *the axes of any trihedral in relation to which the ray m and its focal planes are fixed lie on the paraboloid formed by those transversals of the two rectifying lines which are perpendicular to the ray m.*

When the congruence formed by the rays *m* is a normal congruence, this paraboloid becomes identical with the *paraboloïde des huit droites* of Mannheim.*

---

* Comptes Rendus, vol. 84 (1877), p. 645.

TRINITY COLLEGE,
DUBLIN

ERRATA, VOLUME 31

ARTHUR RANUM, *The singular points of analytic space-curves.*

Page 162, in the table, line 5, third column, for "136" read "135."

Page 162, in the table, between lines 11 and 12, delete the horizontal bar.

Page 163, heading of the 9th column, for "ρ_{zz}" read "$\bar{p}_{zz}$."

Page 163, line 26, 10th column, for "1 or ∞" read "c_1 or ∞."

Page 163, heading of 11th column, for "ρ_{xz}" read "$\bar{p}_{zy}$."

Page 163, heading of 13th column, for "ρ_{yz}" read "$\bar{p}_{yz}$."